

UNIVERZITET U BEOGRADU
FARMACEUTSKI FAKULTET

Nemanja B. Đoković

**SIMULACIJE MOLEKULSKE DINAMIKE,
DIZAJN, SINTEZA I *IN VITRO* ISPITIVANJE
INHIBITORA EPIGENETSKIH PROTEINA KAO
POTENCIJALNIH ANTINEOPLASTIKA**

doktorska disertacija

Beograd, 2023

UNIVERSITY OF BELGRADE
FACULTY OF PHARMACY

Nemanja B. Đoković

**MOLECULAR DYNAMICS SIMULATIONS,
DESIGN, SYNTHESIS AND *IN VITRO*
EVALUATION OF INHIBITORS OF
EPIGENETIC PROTEINS AS POTENTIAL
ANTINEOPLASTIC DRUGS**

Doctoral Dissertation

Belgrade, 2023

Mentor doktorske disertacije:

Prof. dr Katarina Nikolić, redovni profesor
Univerzitet u Beogradu - Farmaceutski fakultet

Članovi komisije:

Doc. dr Slavica Oljačić, docent
Univerzitet u Beogradu – Farmaceutski fakultet

Prof. dr Miloš Petković, vanredni profesor
Univerzitet u Beogradu – Farmaceutski fakultet

Dr Tatjana Srdić-Rajić, naučni savetnik
Institut za onkologiju i radiologiju Srbije

Datum odbrane: _____

„Postoji tajna veza između sporosti i sećanja, između brzine i zaborava.“

Milan Kundera

Zahvalnica

Ova doktorska disertacija je izrađena na Katedri za farmaceutsku hemiju Farmaceutskog fakulteta Univerziteta u Beogradu. Deo istraživanja je sproveden u Laboratoriji za istraživanje sirtuina Univerziteta Istočne Finske, Kuopio, Finska.

Neizmernu zahvalnost dugujem svojoj mentorki, prof. dr Katarini Nikolić, koja me je svojim zalaganjem i velikodušnom podrškom u realizaciji istraživanja oblikovala kao naučnika i bez koje rad na ovoj tezi ne bi bio ni upola uzbudljiv.

Posebno se zahvaljujem svojim dragim koleginicama i kolegama sa Katedre za farmaceutsku hemiju na ukazanom poverenju, velikoj kolegijalnoj i ljudskoj podršci. Naročito se zahvaljujem svom prijatelju i kolegi dr Dušanu Ružiću na nezaboravnim razgovorima i aforizmima koji će zauvek ostati utisnuti između redova ove disertacije.

Želim da se zahvaljujem prof. dr Milošu Petkoviću na neprocenjivim savetima iz organskih sinteza kao i na svim vanredno snimljenim spektrima. Zahvaljujem se i Katedri za organsku hemiju koja je uvek nesebično delila svoje istraživačke resurse.

Zahvaljujem se dr Tatjani Srdić-Rajić na divnoj saradnji kroz koju sam spoznao nove dimenzije istraživanja malignih tumora. Zahvaljujem se i doc. dr Slavici Oljačić na svim konstruktivnim kritikama tokom izrade ove teze.

Veliku zahvalnost dugujem i prof. dr Maiji Lahteli-Kakkonen sa Univerziteta Istočne Finske uz koju su molekulsko modelovanje i in vitro ispitivanja u sezoni belih noći bili jedno posebno i nezaboravno iskustvo. Takođe se zahvaljujem i prof. dr Olalli Vázquez sa Filips Univerziteta u Marburgu na nesebičnom deljenju rezultata iz svoje laboratorije.

Zahvaljujem se svojim najdražim prijateljima koji su mi bili velika moralna podrška.

Najveću zahvalnost dugujem svojoj zauvek najvećoj i najplemenitijoj podršci – mojoj porodici, kojoj i posvećujem ovu disertaciju.

Simulacije molekulske dinamike, dizajn, sinteza i *in vitro* ispitivanje inhibitora epigenetskih proteina kao potencijalnih antineoplastika

SAŽETAK

Među epigenetskim ciljnim mestima lekova, sirtuin 2 (SIRT2) i WD-ponavljajući domen 5 protein (WDR5) se ističu kao dva proteina sa velikim potencijalom za razvoj novih grupa antineoplastika. Uprkos činjenici da su inhibitori ova dva proteina pokazali odlične rezultate u pretkliničkim istraživanjima, nijedan od predstavnika još uvek nije ušao u fazu kliničkih ispitivanja. Razvojem novih protokola za računarsko dizajniranje lekova kroz implementaciju informacija iz velikih baza podataka i/ili molekulske simulacije, kao i njihovom ekperimentalnom validacijom primenom organskih sinteza i *in vitro* bioloških ispitivanja, istraživanja sprovedena u okviru ove teze imala su za cilj da postave osnovu za budući razvoj novih hemijskih grupa inhibitora navedenih epigenetskih proteina kao potencijalnih novih antineoplastika.

U prvoj fazi istraživanja, razvijen je protokol za poboljšano uzorkovanje dinamike vezivnog mesta SIRT2 koji se oslanja na kombinaciju simulacija molekulske dinamike i metadinamičkih simulacija. Primenom razvijenog protokola simulacija, otkriveno je postojanje skrivenog džepa unutar vezivnog mesta SIRT2. Implementacijom konformacionih stanja sa novootkrivenim skrivenim džepom u protokol virtuelnog skrininga, iz do sada neistraženih delova hemijskog prostora inhibitora izdvojena su dva nova potentna i selektivna SIRT2 inhibitora ND18 ($IC_{50} = 58,7 \mu M$) i ND85 ($IC_{50} = 85,9 \mu M$). Daljim ispitivanjem antitumorskog potencijala *hit* jedinjenja ND18 na ćelijskoj liniji adenokarcinoma dojke MDA-MB-231, ND18 je okarakterisano kao necitotoksično jedinjenje koji indukuje zaustavljanje ćelija u G2/M fazi ćelijskog ciklusa i povećava ekspresiju E-kadherina.

U drugoj fazi istraživanja, uspostavljeni su kvantitativni modeli odnosa strukture i aktivnosti, odnosno selektivnost (SIRT1/3) SIRT2 inhibitora. Modeli su trenirani na velikom setu podataka od ukupno 1797 jedinjenja, primenom pet vodećih algoritama mašinskog učenja i četiri različita molekulska prikaza. Najbolji modeli su upotrebljeni za kreiranje *in silico* platforme za dizajn novih SIRT2 inhibitora, nazvane *SIRT2i_Predictor*. U okviru platforme, kreiran je grafički korisnički interfejs uz pomoć koga je moguće efikasno izvesti virtuelni skrining i detaljno ispitati odnose strukture i aktivnosti, odnosno selektivnost nepoznatih jedinjenja. *SIRT2i_Predictor* je dostupan putem linka: https://github.com/echonemanja/SIRT2i_Predictor.

U trećoj fazi istraživanja, izveden je dizajn i sinteza 12 novih SIRT2 inhibitora. Proces dizajna se oslanjao na hipotezu o protein-ligand interakciji *lead* strukture ND18, kao i na predviđanja *SIRT2i_Predictor*-a. Šest sintetisanih jedinjenja je pri koncentraciji od 100 μM pokazalo selektivnu inhibitornu aktivnost veću od 50% što ih čini obećavajućim molekulima za dalja istraživanja. Među sintetisanim jedinjenjima se posebno ističe klasa aril-1,2,3-triazola kao potpuno nove hemijske grupe SIRT2 inhibitora.

U četvrtoj fazi istraživanja izvedena je detaljna karakterizacija protein-ligand interakcije potpuno nove grupe fotoreaktivnih peptidomimetskih inhibitora WDR5. Proračunima slobodne energije vezivanja postavljen je protokol za dizajn novih predstavnika, dok je metadinamičkim simulacijama detaljno ispitan proces fotoizomerizacije. Dodatno, simulacijama je otkriven do sada nepoznati put alosterne komunikacije unutar WDR5 koji može poslužiti kao osnova daljeg dizajna efikasnijih inhibitora.

U petoj fazi istraživanja, oslanjajući se na podatake o osjetljivosti ćelijskih linija adenokarcinoma pankreasa na hemijske probe/lekove, razvijen je bioinformatički protokol za predviđanje sinergističkih kombinacija inhibitora histon-deacetilaza (uključujući sirtuine) i ostalih bioaktivnih molekula. Kroz eksperimentalnu validaciju protokola, unutar sfingolipidnog signalnog puta i povezanih efektorskih puteva, izdvojena su nova ciljna mesta za razvoj novih klasa dualnih inhibitora histon-deacetilaza kao potencijalno efikasnijih terapeutika.

Ključne reči: sirtuin 2, WDR5, racionalni dizajn lekova, mašinsko učenje, molekulska dinamika, metadinamika, virtuelni skrining, antineoplastici, QSAR, bioinformatika

Naučna oblast: Farmacija

Uža naučna oblast: Farmaceutsko-medicinska hemija i strukturna analiza

Molecular dynamics simulations, design, synthesis and *in vitro* evaluation of inhibitors of epigenetic proteins as potential antineoplastic drugs

ABSTRACT

Among the epigenetic drug targets, sirtuin 2 (SIRT2) and WD repeat-containing protein 5 (WDR5) stand out as two protein targets with a great potential for the development of novel antineoplastic drugs. Despite the excellent results in preclinical studies, none of the SIRT2 or WDR5 inhibitors have entered clinical trials. By developing new protocols for computer-aided drug design based on information from large databases and/or molecular simulations, the research conducted within this thesis aimed to establish the foundation for future development of new chemical classes of SIRT2 or WDR5 inhibitors proteins as novel antineoplastic drugs.

In the first phase of the research, a protocol for enhanced sampling of the binding pocket dynamics of SIRT2 relying on metadynamics simulations was developed. By applying the developed protocol, the existence of a cryptic pocket within the SIRT2 binding site was discovered. By incorporating conformational states with the newly discovered cryptic pocket into the virtual screening protocol, two new potent and selective SIRT2 inhibitors, ND18 (IC₅₀ = 58.7 μM) and ND85 (IC₅₀ = 85.9 μM), were discovered from unexplored regions of the chemical space of inhibitors. Further investigation of the anticancer potential of the lead compound ND18 on the breast tumor cell line MDA-MB-231 revealed that ND18, at a concentration of 20 μM, is a non-cytotoxic compound that significantly induces cell cycle arrest at the G₂/M phase and increases the expression of E-cadherin.

In the second phase of the research, quantitative structure-activity relationship (QSAR) models were established to assess the potency, as well as selectivity (SIRT1/3) of novel SIRT2 inhibitors. The models were trained on a large dataset of 1797 compounds using five machine learning algorithms and four different molecular representations. The best-performing models were utilized to create an *in silico* platform for designing new SIRT2 inhibitors, named SIRT2i_Predictor. Developed graphical user interface enable performing virtual screening campaigns and detailed investigation of structure-activity and selectivity relationships of unknown compounds. SIRT2i_Predictor is available at: https://github.com/echonemanja/SIRT2i_Predictor.

In the third phase of the research, the design and synthesis of 12 new SIRT2 inhibitors were performed. The design process relied on the hypothesis of protein-ligand interaction of ND18, as well as the predictions from SIRT2i_Predictor. Six of the synthesized compounds exhibited selective inhibitory activity greater than 50% at a concentration of 100 μM, making them promising molecules for further studies. Among the synthesized compounds, the aryl-1,2,3-triazole derivatives stood out as a completely new group of SIRT2 inhibitors.

In the fourth phase of the research, a detailed characterization of the protein-ligand interaction of a completely new group of light-controlled peptidomimetic inhibitors of WDR5 was performed. The protocol for designing new representatives was established based on binding free energy calculations, while details of photoisomerization were investigated through metadynamics simulations. Simulations revealed a previously unknown pathway of allosteric communication within WDR5, which can serve as a basis for designing more effective inhibitors.

In the fifth phase of the research, relying on data on the sensitivity of pancreatic tumor cell lines to bioactive compounds, a bioinformatics protocol for prediction of synergistic combinations of histone deacetylase inhibitors (including sirtuins) and other bioactive compounds was developed.

Through experimental validation of the protocol, novel targets within the sphingolipid signaling pathway and associated effector pathways, were identified as promising targets for the development of new classes of dual histone deacetylase inhibitors.

Key words: sirtuin 2, WDR5, rational drug design, machine learning, molecular dynamics, metadynamics, virtual screening, antineoplastic, QSAR, bioinformatics

Scientific field: Pharmacy

Scientific subfield: Pharmaceutical-medicinal chemistry and structural analysis

Spisak skraćenica

AI	Veštačka inteligencija
CV	Kolektivna varijabla
DCM	Dihlormetan
DIPEA	<i>N,N</i> -Diizopropiletilamin
DNN	Duboke neuronske mreže
ECFP	eng. <i>Extended Connectivity Fingerprint</i> – ECFP, vrsta fingerprinta
EDCI	<i>N</i> -(3-Dimetilaminopropil)- <i>N'</i> -etilkarbodiimid hidrohlorid
EF	Faktor obogaćenja
HOBt	1-Hidroksibenzotriazol hidrat
KNN	Algoritam k-najbližih suseda
LBVS	Virtuelni skrining zasnovan na poznavanju strukture liganda
MACCS	eng. <i>Molecular ACCess System</i> – MACCS, vrsta fingerprinta
MCC	<i>Matthews</i> -ov korelacioni koeficijent
MD	Molekulska dinamika
MetaD	Dobro temperirana metadinamika
ML	Mašinsko učenje
MLL	eng. <i>Mixed lineage leukemia protein</i> – MLL
MM	Molekulska mehanika
NMR	Nuklearna magnetna rezonanca
PPI	Protein-protein interakcija
PRS	Analiza skeniranja perturbacionog odgovora (eng. <i>Perturbation response scanning</i> – PRS)
QM	Kvantna mehanika
QSAR	Kvantitativni odnos strukture i aktivnosti
RF	Algoritam slučajnih šuma
RMSD	Koren srednjeg kvadratnog odstupanja atomskih pozicija
RMSF	Koren srednje kvadratne fluktuacije atomskih pozicija
ROC	eng. <i>Receiver Operating Characteristics</i> – ROC
ROC_AUC	Površina ispod ROC krive
SBVS	Virtuelni skrining zasnovan na poznavanju 3D-strukture ciljnog mesta
SIRT2	Sirtuin 2
SVM	Algoritam potpunih vektora
TNBC	Trostruko negativni adenokarcinom dojke
VS	Virtuelni skrining
WBM	WDR5-vezujući motiv, oznaka vezivnog mesta na WDR5
WDR5	WD-ponavljajući domen 5 protein
XGBoost	eng. <i>eXtreme Gradient Boosting</i> – XGBoost, implementacija algoritma gradijentnog pojačanja

Sadržaj

1. Uvod	1
1.1. Konformaciona dinamika i molekulske prepoznavanje	2
1.2. Računarske metode u studijama konformacione dinamike proteina.....	3
1.2.1. Molekulska dinamika.....	3
1.2.2. Molekulsko-mehanička polja sile	5
1.2.3. Uzorkovanje retkih događaja simulacijama	7
1.2.4. Metadinamika.....	8
1.3. Racionalni dizajn lekova primenom računarskih metoda	9
1.3.1. Dizajn lekova baziran na poznavanju 3D-strukture ciljnog mesta	10
1.3.2. Dizajn lekova baziran na poznavanju strukture liganda	18
1.3.3. Algoritmi mašinskog učenja u otkriću lekova	21
1.4. Epigenetska ciljna mesta u malignim tumorima	25
1.4.1. SIRT2	28
1.4.2. WDR5	31
1.5. Primena bioinformatičkih analiza u otkriću lekova	34
2. Ciljevi istraživanja	36
3. Materijali i metode.....	38
3.1. Racionalni dizajn SIRT2 inhibitora.....	39
3.1.1. Generisanje parametara polja sile	39
3.1.2. Klasične MD simulacije.....	40
3.1.3. Metadinamička simulacija	41
3.1.4. SBVS.....	41
3.1.5. Građenje i validacija <i>in silico</i> LBVS platforme za otkriće novih SIRT2 inhibitora....	43
3.1.6. Sinteze novih selektivnih SIRT2 inhibitora	46
3.1.7. Biološka testiranja.....	50
3.2. Molekulske modelovanje grupe fotoreaktivnih WDR5 inhibitora.....	52
3.2.1. Modelovanje protein-ligand interakcije linearnih fotoreaktivnih peptidomimetskih inhibitora WDR5	52
3.2.2. Modelovanje protein-ligand interakcije cikličnog fotoreaktivnog peptidomimetskog inhibitora WDR5	54
3.3. Bioinformatička analiza novih ciljnih mesta za razvoj dualnih inhibitora histon-deacetilaza	54
3.3.1. Eksperimentalna validacija sinergizama na MIA Paca-2 i Panc-1 ćelijskim linijama	55
4. Rezultati i diskusija.....	56
4.1. Ispitivanje konformacione dinamike SIRT2 i razvoj protokola za SBDD dizajn novih SIRT2 inhibitora.....	57

4.1.1.	Optimizacije molekulske-mehaničkih parametara polja sile SIRT2 liganada.....	57
4.1.2.	Ispitivanje konformacione dinamike SIRT2 primenom klasične MD simulacije.....	62
4.1.3.	Validacija alternativnih konformacionih stanja primenom retrospektivnog virtuelnog skrininga	65
4.1.4.	Razvoj protokola za poboljšano uzorkovanje konformacionih stanja SIRT2 primenom metadinamičkih simulacija	67
4.1.5.	Implementacija alternativnih konformacionih stanja SIRT2 sa novootkrivenim skrivenim džepom u protokol SBVS	70
4.1.6.	Eksperimentalna validacija multistrukturnog SBVS modela kroz ispitivanje neistraženih delova hemijskog prostora.....	72
4.1.7.	Ispitivanje antitumorskog potencijala <i>hit</i> jedinjenja ND18.....	74
4.2.	Razvoj <i>in silico</i> LBVS platforme za otkriće novih SIRT2 inhibitora	75
4.2.1.	Opis seta podataka upotrebljenog za modelovanje odnosa strukture i aktivnosti SIRT2 inhibitora.....	75
4.2.2.	Razvoj i validacija modela mašinskog učenja	77
4.2.3.	Definisanje okvira <i>SIRT2i_Predictor</i> platforme za otkriće novih SIRT2 inhibitora ...	85
4.2.4.	Studija poređenja modela.....	87
4.3.	Sinteze novih SIRT2 inhibitora na osnovu strukture ND18.....	89
4.4.	Molekulske modelovanje protein-ligand interakcije fotoreaktivnih peptidomimetskih inhibitora WDR5	94
4.4.1.	Molekulske modelovanje protein-ligand interakcije linearnih peptidomimetskih inhibitora WDR5	94
4.4.2.	Molekulske modelovanje protein-ligand interakcije cikličnog peptidomimetskog inhibitora WDR5	98
4.5.	Bioinformatička analiza novih ciljnih mesta za razvoj dualnih inhibitora histon-deacetilaza kao potencijalnih terapeutika u tretmanu adenokarcinoma pankreasa	108
5.	Zaključak	114
6.	Literatura.....	118
7.	Prilozi.....	143
	Prilog 7.1. RMSF analiza trajektorija iz SIRT2 MD simulacija.	144
	Prilog 7.2. Alternativni vezivni modovi liganada (centroidi trajektorija) dobijeni nakon klasičnih MD simulacija SIRT2 sistema.	145
	Prilog 7.3. Rezultati BC analize dobijeni analizom MD simulacija SIRT2.....	146
	Prilog 7.4. Prikaz konvergencije metadinamičkih simulacija SIRT2 sistema.	149
	Prilog 7.5. Supstance testirane nakon prospektivnog SBVS.....	151
	Prilog 7.6. Poređenje različitih SBVS modela.	152
	Prilog 7.7. Optimizacija hiperparametara modela mašinskog učenja.	152
	Prilog 7.8. Rezultati interne i eksterne validacije modela mašinskog učenja.	155
	Prilog 7.9. Rezultati eksperimentalne validacije <i>SIRT2i_Predictor</i> -a.	164

Prilog 7.10. NMR (^1H i ^{13}C) spektri prekursora i finalnih jedinjenja.	166
Prilog 7.11. Modelovane strukture kompleksa WDR5 sa lineranim fotoreaktivnim peptidomimeticima.	182
Prilog 7.12. Rezultai analize dinamičkih mreža interakcija rezidua WDR5.....	185

1. Uvod

1.1. Konformaciona dinamika i molekulska prepoznavanje

Slavni fizičar Ričard Fajnman je još 1963. godine izjavio da se sve što živa bića čine može objasniti kroz pokrete atoma. Od ove čuvene izjave pa do danas, izučavanje dinamike pokreta atoma unutar biomakromolekula (u daljem tekstu – konformaciona dinamika) postalo je jedan od osnovnih stubova savremenih biomedicinskih istraživanja. Počevši od termalnih fluktuacija gradivnih jedinica biomakromolekula pa sve do većih konformacionih promena unutar njihovih strukturnih celina, pokreti unutar biomakromolekula predstavljaju osnovu njihove funkcije^{1,2}.

Konformacione dinamika biomakromolekula predstavlja sastavni deo procesa molekuskog prepoznavanja koji se nalazi u osnovi funkcionisanja bioloških sistema. Termin molekulska prepoznavanje se odnosi na specifično prepoznavanje i vezivanje dva ili više molekula putem nekovalentnih interakcija. Najčešći tipovi nekovalentnih interakcija koji su uključeni u molekulska prepoznavanje su elektrostatičke interakcije, vodonične veze, pi-pi i *van der Waals*-ove interakcije. Protein-ligand, protein-protein i protein-DNK interakcije su primeri molekuskog prepoznavanja u biološkim sistemima. Ove interakcije su od suštinskog značaja za mnoge ćelijske procese, uključujući prenos signala, regulaciju ekspresije gena, enzimska katalizu itd. Specifičnost i trajanje ovih interakcija su dva glavna aspekta molekuskog prepoznavanja koja su ključna za biološku funkciju³⁻⁵.

Specifičnost molekuskog prepoznavanja proizlazi iz komplementarnosti interaktivnih površina molekula. Pri protein-ligand interakcijama, ligand se vezuje za specifično vezno mesto na proteinu komplementarnog oblika i hemijskog okruženja. Specifičnost interakcije takođe je definisana jačinom i rasporedom nekovalentnih veza koje se formiraju između dva interakciona partnera^{5,6}. Razumevanje principa molekuskog prepoznavanja, posebno fizičkih sila koje upravljaju afinitetom i specifičnošću, je od velikog značaja u biomedicinskim istraživanjima i otkriću lekova. Izučavanje konformacione dinamike ključnih aktera molekuskog prepoznavanja predstavlja jedan od osnovnih aspekata savremenog racionalnog dizajna lekova. Da bi se razumela konformaciona dinamika biomolekula, razvijene su različite eksperimentalne i računarske tehnike. Kristalografija X-zraka i nuklearna magnetno-rezonantna (NMR) spektroskopija mogu se koristiti za određivanje trodimenzionalne strukture biomolekula u različitim konformacijama, dok se fluorescentno-rezonantni energetski transfer (FRET) i druge spektroskopske tehnike mogu koristiti za praćenje promena u međumolekulske udaljenosti i orijentaciji^{7,8}. Kristalografija X-zracima danas predstavlja jednu od najčešće korišćenih eksperimentalnih metoda za izučavanje strukture biomakromolekula na atomskoj rezoluciji. Ipak, ovom metodom se dobijaju informacije o vrlo ograničenom setu konformacionih stanja koji se mogu zateći unutar kristala što najčešće ne odražava u potpunosti konformaciono ponašanje u rastvoru⁹⁻¹².

U idealnom slučaju, metoda za praćenje konformacione dinamike treba da omogući praćenje molekula, uz punu atomsku rezoluciju na vremenskoj skali od najmanje jedne mikrosekunde. Jedino metode računarskih simulacija ispunjavaju sve navedene uslove⁹. Uprkos ograničenjima ovih metoda proisteklim iz određenih aproksimacija uvedenih u računarske proračune, metode računarskih simulacija su uznapredovale do nivoa na kom se smatraju komplementarnim eksperimentalnim metodama zbog čega se često oslovljavaju i terminom „računarski mikroskop“¹⁰. Uz ubrzani hardverski i softverski razvoj poslednjih decenija, metode računarskih simulacija konformacione dinamike su ostvarile veliki napredak u tačnosti i preciznosti. Danas se ove metode mogu upotrebiti za izučavanje dinamike procesa koji se odigravaju na vremenskoj skali od nanosekunde do milisekunde. Jedna od metoda koja predstavlja zlatni standard u izučavanju konformacione dinamike proteina jeste i molekulska dinamika^{9,13}.

1.2. Računarske metode u studijama konformacione dinamike proteina

1.2.1. Molekulska dinamika

Molekulska dinamika (MD) je računarska metoda za simulaciju kretanja atoma unutar molekula. Metoda je bazirana na zakonima klasične mehanike i koristi skup jednačina kojima se opisuje kretanje svakog atoma u sistemu. Rezultirajuća trajektorija atoma pruža informacije o termodinamičkim i kinetičkim svojstvima sistema. Jedna od glavnih prednosti MD simulacija je omogućavanje uvida u dinamiku i ponašanje molekulskih sistema koji su teški ili nemogući za eksperimentalno praćenje^{2,14}.

U osnovi molekulske dinamike stoje algoritmi za rešavanje jednačine kretanja atoma u simuliranom sistemu koji uzimaju u obzir njihove međusobne interakcije¹⁵. Sam proračun se oslanja na *Newton*-ove zakone kretanja koji silu (F_i) koja deluje na i -tu česticu u vreme t izražavaju kroz proizvod mase (m_i) i ubrzanja (a_i) i -te čestice (jednačina 1).

$$F_i = m_i a_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad (1)$$

gde je $r_i(t) = (x_i(t), y_i(t), z_i(t))$ pozicioni vektor i -te čestice. Sila koja deluje na i -tu česticu može se izraziti i kao parcijalni izvod potencijalne energije (V) po odgovarajućoj koordinati čestice (jednačina 2).

$$F_i = - \frac{\partial V}{\partial r_i} \quad (2)$$

Potencijalna energija interakcije je moguće izračunati primenom molekulske-mehaničkih polja sile. Polja sile predstavljaju setove empirijski određenih energetske funkcije i parametara kojima se opisuje zavisnost potencijalne energije od koordinate čestice u simuliranom sistemu (detaljnije u poglavlju 1.2.2)¹⁶. Dakle, iz potencijalne energije svake čestice (atoma) moguće je izračunati pojedinačno ubrzanje svake od čestica unutar simuliranog sistema što bi teorijski rešilo navedenu jednačinu kretanja. Usled velikog broja čestica, jednačinu kretanja je nemoguće rešiti analitički, već se primenjuju numeričke metode. Numeričkim rešenjem jednačine kretanja, kretanje se razlaže na veliki broj malih subsekventnih koraka u fiksnim vremenskim intervalima (δt)^{15,17,18}. Na taj način se pozicija i -te čestice u vremenu $t + \delta t$ može aproksimirati primenom *Taylor*-ove serije kao:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \dots \quad (3)$$

gde su v brzina (prvi izvod pozicije u odnosu na vreme), a ubrzanje (drugi izvod pozicije čestice u odnosu na vreme).

Slično važi i za poziciju čestice u vremenu $t - \delta t$:

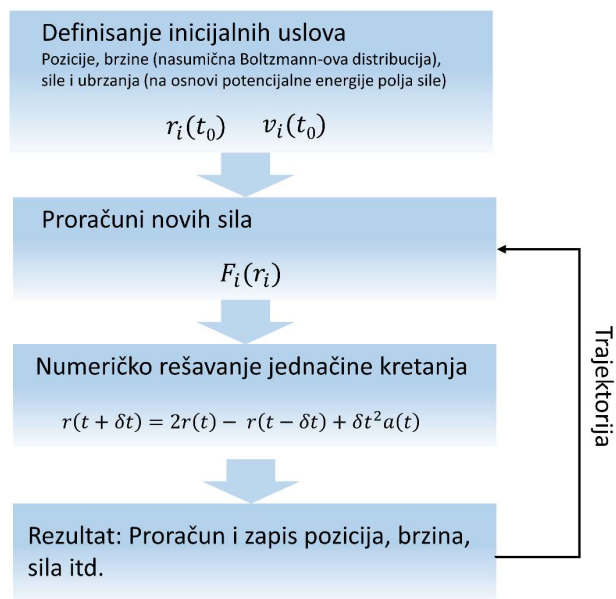
$$r(t - \delta t) = r(t) - \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \dots \quad (4)$$

Sabiranjem jednačina (3) i (4), dobija se glavni izraz tzv. *Verlet*-ove jednačine za rešavanje pozicije atoma (jednačina 5).

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \quad (5)$$

Verlet-ov algoritam predstavlja jedan od najpoznatijih algoritama za integraciju jednačine kretanja. Oslanja se na poznavanje pozicije i /ili ubrzanja u vremenu t i $t - \delta t$ kako bi izračunao poziciju čestice u vremenu $t + \delta t$. Danas se koriste napredniji algoritmi izvedeni iz klasičnog *Verlet*-ovog algoritma, poput *Velocity Verlet* i *leapfrog* algoritma koji generišu manju grešku u proračunu trajektorije atoma (trajektorija – Kartezijanove koordinate atoma tokom vremena).^{19,20}

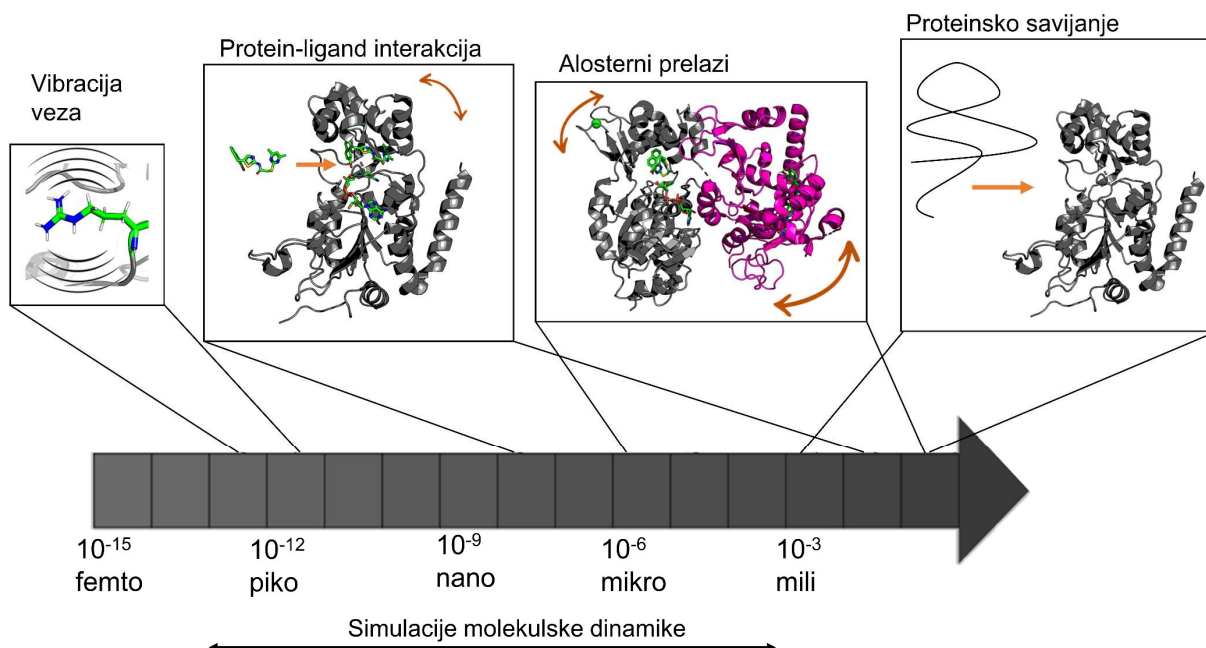
Implementacija *Verlet*-ovog i srodnih algoritama u programe za MD simulacije je jednostavna i prati pojednostavljenu šemu prikazanu na Slici 1.1²¹.



Slika 1.1. Prikaz osnovnog MD algoritma.

MD simulacije se mogu koristiti za proučavanje različitih molekulskih sistema, kao što su proteini, nukleinske kiseline, ugljeni hidrati i lipidi. Mogu se koristiti i za proučavanje materijala, poput polimera i nanočestica. Krajnji cilj MD simulacija jeste predviđanje makroskopskih termodinamičkih osobina sistema (afinitet liganda za protein, temperatura, gustina, pritisak itd.) izvedenih iz mikroskopskih stanja sistema prema zakonima statističke termodinamike. Prema osnovnim principima statističke termodinamike, makroskopska osobina može biti izračunata iz srednjih vrednosti gustina verovatnoće nalaženja sistema u određenoj mikroskopskoj konfiguraciji (poglavlje 1.3.1.d) pod pretpostavkom ergodičnosti simuliranog sistema²². U sistemu koji je ergodičan, vremenski računata prosečne osobine sistema, kao što su prosečna energija ili konformaciona distribucija, biće jednake njihovim prosečnim vrednostima za čitav ansambl, odnosno osobinama izračunatim na osnovu srednjih vrednosti svih mogućih konformacija koje sistem može da zauzme. U ergodičnim sistemima, sistem će posetiti svako dostupno konformaciono stanje, bez obzira na njegovo početno stanje²². Međutim, ergodičnost sistema se može postići samo izuzetno dugom MD simulacijom koja zahteva veliki utrošak računarskih resursa što predstavlja jedno od glavnih ograničenja ove metode. Ograničenja u računarskim resursima još uvek sprečavaju klasične MD simulacije da dostignu simulacije konformacione dinamike procesa koji se dešavaju na vremenskoj skali od milisekunde (Slika 1.2)^{13,22-24}.

Drugo veliko ograničenje MD predstavlja i tačnost polja sile koje se koristi za modelovanje sistema. S obzirom da polja sile samo aproksimiraju kvantno-mehaničke procese, nemoguće ih je upotrebiti za izučavanje sistema u kojima kvantni procesi igraju bitnu ulogu, kao što su procesi u koje su uključeni prelazni metali, ili procesi u kojima dolazi do raskidanja kovalentnih veza. Ipak, treba istaći da su danas usavršene i metode kojima je moguće u velikoj meri prevazići navedena ograničenja. Tako je ograničenje u vremenskoj skali klasičnih MD simulacija moguće prevazići primenom tzv. metoda poboljšanog uzorkovanja, dok je dinamiku sistema u kojima kvantni procesi igraju važnu ulogu moguće izučavati hibridnim kvantno-mehaničkim/molekulsko-mehaničkim (QM/MM) pristupom^{25,26}.



Slika 1.2. Ograničenje u vremenskoj skali klasičnih MD simulacija.

1.2.2. Molekulsko-mehanička polja sile

Računarske metode koje se koriste za simulacije molekulskih sistema se oslanjaju na dva fundamentalna pristupa: kvantno-mehanički (QM) i molekulsko-mehanički (MM) pristup. Primenom QM pristupa se uzimaju u obzir karakteristike elektrona i jezgra atoma usled čega je ovim pristupom moguće izvesti veoma precizne proračune. Oslanjanjem na savremene aproksimacije za rešavanje *Schrödinger*-ove jednačine, QM metode mogu da dostignu tačnost veću od 99%, odnosno grešku na nivou 1 kcal/mol.²⁷ Ipak, usled visokih zahteva za računarskim resursima, ove metode još uvek nije moguće primeniti na velike molekulske sisteme poput biomakromolekula, te glavna primena ovih metoda ostaje na nivou izučavanja malih molekula i njihovih osobina. Sa druge strane, MM pristup koristi koncepte klasične mehanike za modelovanje i simulaciju ponašanja molekula. U osnovi, molekulska mehanika tretira molekule kao skup čestica koje se ponašaju u skladu sa *Newton*-ovim zakonima kretanja. Glavna aproksimacija MM pristupa predstavlja posmatranje molekula kao skupa čestica (atoma) fiksiranog naelektrisanja i energije koja je raščlanjena na doprinose veza, uglova i nevezivnih interakcija. Funkcija potencijalne energije kojom se aproksimira energija sistema u odgovarajućoj konfiguraciji nosi generalni naziv polje sile. Ovom aproksimacijom značajno se smanjuje potreba za računarskim resursima zbog čega MM metode imaju širu primenu u izučavanju biomakromolekula. Uz konstantna usavršavanja polja sile, danas se primenom ovih metoda postiže visok nivo tačnosti proračuna.^{18,28,29}

Polja sile se sastoje od nekoliko delova koji opisuju različite vrste interakcija između atoma i molekula. Vezivne interakcije, kao što su protezanje veze, savijanje ugla veze i torziona rotacija, opisuju interakcije između vezanih atoma i obično se modeluju harmonijskim ili kosinusnim funkcijama. Nevezivne interakcije, kao što su *van der Waals*-ove interakcije i elektrostatičke *Coulomb*-ove interakcije, se modeluju empirijskim jednačinama koje uzimaju u obzir zavisnost ovih interakcija od međusobne udaljenosti.^{15,18}

Opšta formula za izračunavanje potencijalne energije primenom polja sile obično ima sledeću formu (jednačina 6):

$$V_{total} = V_{veza} + V_{ugao} + V_{diedar} + V_{vdW} + V_{elec} \quad (6)$$

V_{veza} , V_{ugao} i V_{diedar} predstavljaju delove jednačine kojima se opisuje doprinos intramolekulskih osobina molekula nastalih istežanjem kovalentnih veza, savijanjem uglova veza, ili rotacijom oko diedarskih uglova. Ova tri parametra se računaju na sledeći način (jednačine 7-9):

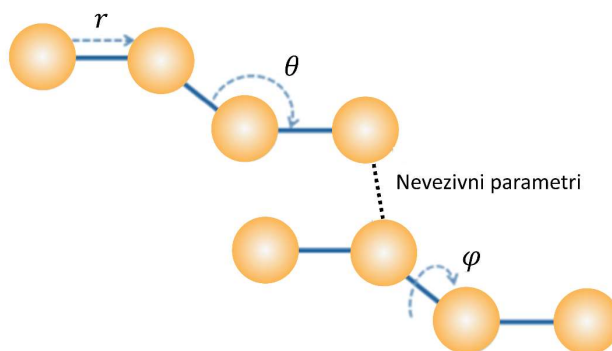
$$V_{veza} = \sum_{veze} \frac{1}{2} k_b (r - r_0)^2 \quad (7)$$

$$V_{ugao} = \sum_{uglovi} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (8)$$

gde k_b i k_θ predstavljaju konstante sile, a r_0 i θ_0 ravnotežne vrednosti dužine veze (r) i ugla veze (θ). V_{veza} i V_{ugao} imaju formulaciju potencijalne energije klasičnog harmonijskog oscilatora. Doprinos diedarske energije se računa iz rotacije veze unutar četvorokomponentnog fragmenta sastavljenog od atoma povezanih sa tri kovalentne veze (Slika 1.3). Energetske barijere pri ovoj rotaciji se predstavljaju jednačnom:

$$V_{diedar} = \sum_{diedri} \frac{1}{2} k_\varphi [1 + \cos(n\varphi - \delta)] \quad (9)$$

gde je φ diedarski ugao, k_φ konstanta sile, n multiplicitet i δ faza. Multiplicitet predstavlja broj ciklusa pri rotaciji od 360° diedarskog ugla, dok faza ukazuje na lokaciju minimuma energije usled nepovoljnih sternih interakcija preostalih atoma pri rotaciji veze u četvoroatomnom fragmentu.



Slika 1.3. Prikaz četvoroatomnih fragmenata sa vezivnim i nevezivnim interakcija u molekulsko-mehaničkom polju sile.

Nevezivni parametri funkcije potencijalne energije su podeljeni na Lennard-Jones 12-6 potencijal kojim se izražavaju *van der Waals*-ove interakcije i *Coulomb*-ov potencijal za elektrostatske interakcije (jednačine 10 i 11). *Lennard-Jones*-ova funkcija se sastoji od parametra koji opisuje repulzije kratkog dometa između dva atoma na rastojanju r_{ij} što predstavlja model *Pauli*-jeve repulzije usled preklapanja elektronskih orbitala ($\left(\frac{1}{r_{ij}}\right)^{12}$). Drugi parametar *Lennard-Jones* jednačine jeste parametar koji opisuje privlačne *van der Waals*-ove interakcije odnosno disperzione sile između permanentnih i indukovanih dipola ($\left(\frac{1}{r_{ij}}\right)^6$).

$$V_{vdW} = \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] \quad (10)$$

ϵ_{ij} predstavlja dubinu potencijalne jame interakcije para atoma, dok je σ distanca između dva atoma na kojoj je interakciona energija jednaka nuli.

$$V_{elec} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} \quad (11)$$

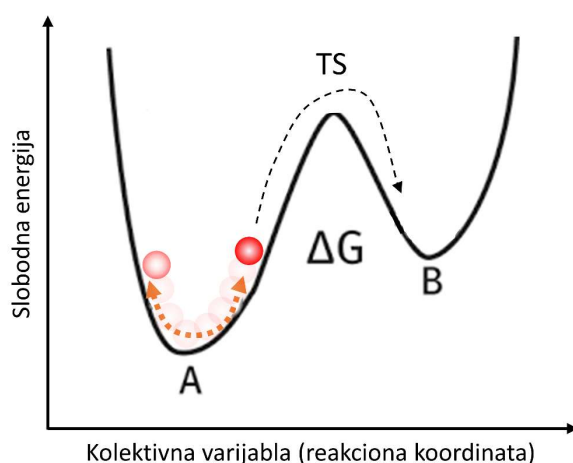
Coulomb-ov potencijal (jednačina 11) predstavlja model repulzionih i atraktivnih elektrostatskih interakcija između atoma izučavanog sistema, gde su q_i i q_j parcijalna naelektrisanja atoma, dok je r_{ij} distanca između njih. ϵ i ϵ_0 predstavljaju relativnu dielektričnu konstantu i permitivnost vakuuma.

Parametri u polju sila se obično dobijaju iz eksperimentalnih podataka ili QM proračuna. Polje sila se zatim kalibriše prilagođavanjem parametara da što preciznije reprodukuju eksperimentalne podatke ili QM proračune. Nakon kalibracije, polja sile se mogu koristiti za simuliranje ponašanja složenih molekulskih sistema koji su teški ili nemogući za eksperimentalno proučavanje. Parametri polja sile se uglavnom definišu za određene strukturne celine što otvara mogućnost njihovog skaliranja na veliki broj različitih molekula. Na primer, parametri definisani za aminokiseline se mogu upotrebiti za sve izučavane proteine jer se očekuje da sama priroda proteina ne utiče u velikoj meri na ponašanje aminokiseline u svom okruženju. Situacija je kompleksnija kada su u pitanju mali organski molekuli gde postoji veliki diverzitet u hemijskim strukturama. Problem parametrizacije malih molekula je danas delimično rešen građenjem baza podataka iz kojih se mogu preuzeti parametri za strukturno slične gradivne jedinice. Međutim, neretko se dešava da za neke delove struktura ne postoji dovoljno sličan fragment u bazama podataka zbog čega se, radi postizanja veće preciznosti pri simulacijama, mora pristupiti izvođenju parametara iz QM proračuna.^{25,30,31}

1.2.3. Uzorkovanje retkih događaja simulacijama

Statističkom mehanikom je moguće povezati mikroskopske karakteristike simuliranog sistema sa makroskopskim veličinama koristeći verovatnoću nalaženja sistema u odgovarajućoj konfiguraciji definisanoj položajem N atoma i njihovim momentima. Prostor od $6N$ dimenzija definisanih momentumom i pozicijom čestica se drugačije naziva fazni prostor.²²

Na osnovu teorije prelaznih stanja, mnogi konformacioni prelazi u molekularnim sistemima dešavaju se preko visokoenergetskih prelaznih stanja. U skladu sa *Boltzmann*-ovom distribucijom, ova visokoenergetska stanja prati i eksponencijalni pad u verovatnoći uzorkovanja primenom klasične molekulske dinamike u realnim vremenskim okvirima simulacija. Da bi prešao iz jednog stanja minimuma energije u drugo, sistem mora da poseti pojedine visokoenergetske konfiguracije (Slika 1.4). Prema ovoj teoriji, suština problema uzorkovanja retkih događaja leži u identifikaciji važnih stepeni slobode koji karakterišu mehanizam prelaza (tzv. reakcionih koordinata).²²



Slika 1.4. Teorija prelaznog stanja i verovatnoća prelaska sistema (crvena sfera) iz stanja A u stanje B preko visokoenergetskog prelaznog stanja (TS). Narandžastom isprekidanom linijom je označen deo faznog prostora dostupan za istraživanje klasičnim MD simulacijama. Isprekidanom sivom linijom je naznačena mala verovatnoća prelaska energetske barijere klasičnom MD simulacijom.

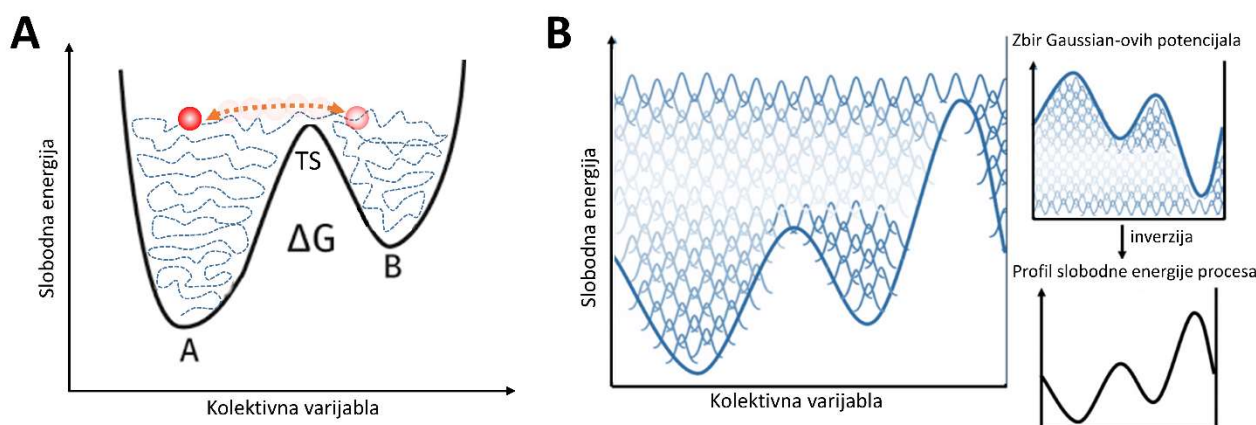
Visoka dimenzionalnost velikih konformacionih prelaza u molekulima čini vrlo teškim analizu uloge svih stepeni slobode u faznom prostoru reaktivnog događaja. Uopšteno govoreći, visoka dimenzionalnost faznog prostora predstavlja jedan od glavnih razloga zašto simulacijama klasične molekulske dinamike, uz racionalan utrošak računarskih resursa, nije moguće adekvatno istražiti sva moguća konfiguraciona stanja biomakromolekula^{22,32,33}.

Teorija prelaznog stanja pruža jednostavan teorijski okvir koji je postao osnova za razvoj računarskih tehnika za proučavanje događaja prelaska preko visokoenergetskih barijera. Kao što je već spomenuto, verovatnoća uzorkovanja visokoenergetskih prelaznih stanja eksponencijalno opada sa visinom energetskih barijera³³. Da bi se rešio problem eksponencijalnosti vremenske skale, razvijene su tehnike koje poboljšavaju uzorkovanje unutar faznog prostora. Ove metode uzorkovanja retkih događaja mogu se grubo podeliti u tri kategorije: 1) metode koje poboljšavaju uzorkovanje povećanjem temperature; 2) metode koje primenjuju tzv. potencijal pristrasnosti na jedan ili više izabranih kolektivnih varijabli (CV) i 3) metode koje pronalaze prelazne putanje između dve stabilne faze. Svaka metoda ima svoje prednosti i mane, zbog čega odabir prave metode zavisi dosta od konteksta simulacija.³⁴

1.2.4. Metadinamika

Jedan način da se reši problem uzorkovanja retkih događaja jeste projekcija visokodimenzionalnog konfiguracionog prostora u nekoliko geometrijskih parametara koji su dobri kandidati za opisivanje energetskog prelaza od interesa. Ove parametre nazivamo kolektivnim varijablama (CV).^{22,35}

Kolektivne varijable su funkcije konfiguracionog prostora koje omogućavaju redukciju višedimenzionalnog faznog prostora na svega nekoliko stepeni slobode. Primeri kolektivnih varijabli koje se koriste u izučavanju retkih događaja mogu biti jednostavni parametri, kao što su dužine veza, uglovi, torzioni uglovi itd., ili u slučaju složenijih tranzicija, mogu biti manje jednostavni parametri, kao što su koren srednjeg kvadratnog odstupanja (eng. *Root Mean Square Deviation* – RMSD) od koordinata datog referentnog konformacionog stanja, koordinacioni brojevi, dipolni moment itd. Odabir CV zavisi od konkretnog sistema koji se proučava i istraživačkog pitanja koje se razmatra.³⁵



Slika 1.5. Princip rada metadinamike. A – Putanja (isprekidana linija) sistema (crvena sfera) između dva energetska minimuma prelaskom energetske barijere prelaznog stanja (TS); B – Deponovanje potencijala pristrasnosti tokom metadinamičkih simulacija i konverzija deponovanih potencijala u profil slobodne energije. Prilagođeno na osnovu reference³⁶.

Metadinamika je računarska tehnika koju su razvili Laio i Parrinello radi poboljšanja uzorkovanja retkih događaja i proučavanja slobodne energije kompleksnih molekulske sistema.^{37,38} Metoda se zasniva na obeshrabrivanju sistema da poseti već posećena stanja u projekciji faznog prostora na unapred definisane kolektivne varijable. Metadinamika se zasniva na principu dodavanja tzv.

istorijski zavisnog potencijala pristrasnosti na funkciju potencijalne energije sistema projektovanog na prostor definisan kolektivnim varijablama. Postepeno dodavanje ovog potencijala dovodi do povećavanja energije sistema, što omogućava prelazak energetskih barijera definisanih u CV prostoru i istraživanje prethodno nedostupnih stanja (Slika 1.5).

Kada se istraži celokupna površina slobodne energije, dolazi do slobodne difuzije sistema u okviru CV prostora, odnosno do konvergiranja simulacije. Pokazano je da je ukupni primenjeni potencijal pristrasnosti u direktnoj vezi sa negativnom vrednošću slobodne energije izučavanog prelaza zbog čega se metadinamika može koristiti i u kvantifikaciji energetike određenog retkog događaja (Slika 1.5)^{39,40}.

Kao ishod simulacija, metadinamika može pomoći u otkrivanju novih reakcionih puteva, utvrđivanju termodinamike i kinetike reakcija i karakterizaciji konformacijske dinamike molekula. U krajnjem slučaju, na ovaj način se prevazilazi osnovni problem uzorkovanja simulacijama molekulske dinamike, koje često ostaju zaglavljene u konfiguracionim stanjima lokalnih minimuma (Slika 1.4)³⁵.

Potencijal pristrasnosti (V_{bias}) se može predstaviti jednačinom 12:

$$V_{bias}(s, t) = \omega \sum_{t' < t} \exp\left(-\frac{(s_{t'} - s)^2}{2\sigma^2}\right) \quad (12)$$

gde je s vrednost CV, ω visina *Gaussian*-ove funkcije potencijala pristrasnosti, σ širina funkcije, t vreme odnosno faktor povezan sa brzinom deponovanja potencijala. U klasičnoj verziji metadinamike, parametri poput visine, širine i brzine deponovanja *Gaussian*-ovih potencijala pristrasnosti definišu se pre početka simulacije. Dodavanje istih *Gaussian*-a na početku simulacije kao i na kraju dovodi do oscilacije sistema oko ravnotežnog položaja, odnosno prevenira konvergenciju simulacije⁴¹. Novija adaptacija metadinamike podrazumeva dobro-temperovanu metadinamiku u kojoj se visina *Gaussian*-ovog potencijala postepeno smanjuje tokom simulacije čime omogućava efikasniju konvergenciju³⁸. Metadinamika je primenjiva na širok spektar molekularnih sistema, uključujući proteine, nukleinske kiseline i male molekule. Metadinamikom je moguće proučavati vezivanja proteina i liganda, protein-protein interakcija, difuziju kroz ćelijsku membranu itd.

Iako efikasna, metadinamika ima i određena ograničenja. Pravilan odabir kolektivnih varijabli predstavlja kritičan korak u metadinamičkoj simulaciji. Deponovanje potencijala pristrasnosti na neadekvatne CV može da odgura sistem u nefizička stanja i da pruži iskrivljenu sliku o izučavanim procesima. Ukoliko se pri projekciji sistema na odabrana/e CV ne izdvajaju jasno visokoenergetska stanja, postoji velika verovatnoća da uzorkovanje neće, ili će samo delimično, biti poboljšano zahtevajući velike količine računarskih resursa i dugotrajna vremena simulacije³⁵.

1.3. Racionalni dizajn lekova primenom računarskih metoda

Proces otkrića novih lekova se smatra dugotrajnim, skupim i visoko rizičnim procesom. U proseku protekne između 10 i 15 godina i utroši se 2.8 milijardi USD do izlaska novog leka na tržište dok je visok rizik od neuspeha nezaobilazan. Sve strožiji zahtevi regulatornih tela kao i rast u ceni kliničkih studija značajno povećavaju krajnju cenu novih lekova što motiviše farmaceutske kompanije da se sve više okreću inovativnim načinima otkrića novih lekova. Primenom inovativnih metoda skraćuje se put do novog leka, smanjuju se troškovi i u krajnjoj liniji čuva se finansijska održivost ovog kompleksnog procesa.^{42,43}

Primena inovativnih rešenja, koja u velikoj meri podrazumevaju primenu naprednih računarskih metoda otkrića lekova, pokazala se kao opravdana i korisna strategija. U prilog ovoj tvrdnji ide i činjenica da je 2018. najveći broj lekova potekao upravo od malih biotehnoških kompanija od kojih su neke osnovane i pri univerzitetima. Uprkos velikim razlikama u sredstvima i resursima kojima raspolazu male biotehnoške kompanije i industrija, ovaj napredak se može objasniti

racionalnijim pristupom u dizajnu novih lekova koji je efikasniji i ekonomičniji u odnosu na tradicionalne metode otkrića lekova primenjivane od strane farmaceutske industrije^{43,44}.

Racionalni dizajn lekova je složen multidisciplinarni proces koji uključuje identifikaciju farmakološki aktivnih molekula, optimizaciju njihove farmakološke aktivnosti, toksičnosti i farmakokinetike, a sve u cilju razvoja efikasnih i bezbednijih lekova. Dva glavna pristupa su dizajn lekova baziran na poznavanju 3D-strukture ciljnog mesta (eng. *Structure-based Drug Design* – SBDD) i dizajn lekova baziran na poznavanju strukture liganda (eng. *Ligand-based Drug Design* – LBDD).⁴⁵

1.3.1. Dizajn lekova baziran na poznavanju 3D-strukture ciljnog mesta

SBDD pristup se oslanja na strukturne informacije o ciljnom proteinu, odnosno na poznavanje njegove trodimenzionalne strukture, čime se omogućava dizajniranje molekula koji se selektivno vezuju za vezivno mesto proteina. Tradicionalni protokol SBDD započinje definisanjem 3D-strukture vezivnog mesta. Ovaj korak se u velikoj meri oslanja na metode strukturne biologije poput kristalografije X-zraka, ili krio-elektornske mikroskopije, na neke od metoda molekuskog modelovanja poput homologog modelovanja, a u novije vreme i na modele dubokih neuronskih mreža za predviđanje 3D-strukture proteina.⁴⁶ Nakon definisanja vezivnog mesta, primena odgovarajućih računarskih metoda je neizostavna u identifikaciji potpuno novih grupa liganada koji se vezuju za definisano vezivno mesto, ili u optimizaciji postojećih liganada. U najčešće korišćene metode SBDD spadaju molekulski doking, strukturno-zavisni virtuelni skrining i simulacije molekulske dinamike. Proteini do danas ostaju najvažnija ciljna mesta dejstva lekova, zbog čega izučavanje molekuskog prepoznavanja između proteina i malog molekula, odnosno protein-ligand interakcije, predstavlja jednu od ključnih faza u racionalnom dizajnu lekova zasnovanog na poznavanju 3D-strukture ciljnog mesta.⁴⁷

1.3.1.a. Opšta termodinamička razmatranja protein-ligand interakcije

Interakcijom proteina i liganda dolazi do formiranja reverzibilnog kompleksa usled uspostavljanja nekovalentnih intermolekulskih interakcija. Ova interakcija može da rezultuje u različitim promenama u aktivnosti proteina, koje su u velikom broju slučajeva tesno povezane sa konformacionim promenama unutar samog proteina.^{3,48}

Racionalizacija protein-ligand interakcija igra važnu ulogu u dizajnu lekova. Nakon identifikacije proteina koji ima važnu ulogu u patofiziologiji bolesti, uobičajen proces prekliničkog razvoja lekova podrazumeva dizajn serije liganada koji su sposobni da interaguju sa ciljnim mestom i modifikuju njegovu funkciju. U ovom procesu, poznavanje strukturnih detalja, termodinamičkih kao i kinetičkih osobina interakcija može usmeriti dizajn ka postizanju visokoafinitetnog i selektivnog vezivanja liganda što se smatra ultimativnim ciljem SBDD pristupa.

Ravnotežna reakcija proteina (P) i liganda (L) se može predstaviti jednačinom 13:



gde su k_{on} i k_{off} konstante brzine reakcije vezivanja, odnosno disocijacije liganda i predstavljaju veličine kojima se definiše kinetika protein-ligand interakcije. Njihovim količnikom definiše se termodinamička veličina kojom se opisuje afinitet liganda ka svom ciljnom mestu – konstanta vezivanja liganda (K_b), izražena jednačinom:

$$K_b = \frac{k_{on}}{k_{off}} = \frac{[PL]}{[P][L]} = \frac{1}{K_d} \quad (14)$$

gde je K_d konstanta disocijacije liganda.

Recipročna vrednost k_{off} ($1/k_{off}$) se često oslovljava kao vreme zadržavanja liganda u ciljnom mestu. U poslednje vreme se sve više pažnje posvećuje optimizaciji ove veličine pri dizajnu lekova usled bolje korelacije sa aktivnostima u *in vivo* uslovima u odnosu na ravnotežnu konstantu vezivanja liganda (K_b). Naime, *in vivo* uslovi uglavnom podrazumevaju otvoren sistem, nasuprot uslovima zatvorenog termodinamičkog sistema u kojima se određuje ravnotežna konstanta vezivanja zbog čega mogu postojati odgovarajuća odstupanja kada se ligand dizajniran na osnovu K_b testira *in vivo*. Ipak, sa aspekta metoda računarskog dizajna, praćenje kinetike interakcije zahteva veliko računarsko vreme zbog čega ova veličina neće biti dalje razmatrana.^{6,6,49}

Sa aspekta termodinamičkih razmatranja, protein-ligand interakcija se može predstaviti kroz promenu *Gibbs*-ove slobodne energije procesa vezivanja (ΔG). Analogno sa svim spontanim procesima, proces vezivanja liganda za svoj protein dešava se samo u slučaju kada je promena u *Gibbs*-ovoj slobodnoj energiji između dva stanja negativna vrednost, posmatrano u ravnotežnim uslovima pri konstantnom pritisku i temperaturi. Vrednost ΔG je u direktnoj vezi sa afinitetom liganda za odgovarajući protein kroz relaciju 15:

$$\Delta G = -RT \ln K_b \quad (15)$$

gde je R univerzalna gasna konstanta, a T temperatura.

Dalje, ΔG se može razložiti na doprinose entalpijskih (ΔH) i entropijskih promena (ΔS):

$$\Delta G = \Delta H - T\Delta S \quad (16)$$

Promena u entalpiji tokom procesa interakcije proteina i liganda reflektuje energetske promene nastale usled raskidanja i uspostavljanja različitih nekovalentnih veza tokom samog procesa vezivanja. Ova promena uključuje i promene u interakciji između molekula rastvarača sa proteinom i/ili ligandom nastalih usled vezivanja liganda za svoj protein. Promena u entropiji sistema tokom procesa protein-ligand interakcije se može podeliti na različite doprinose:

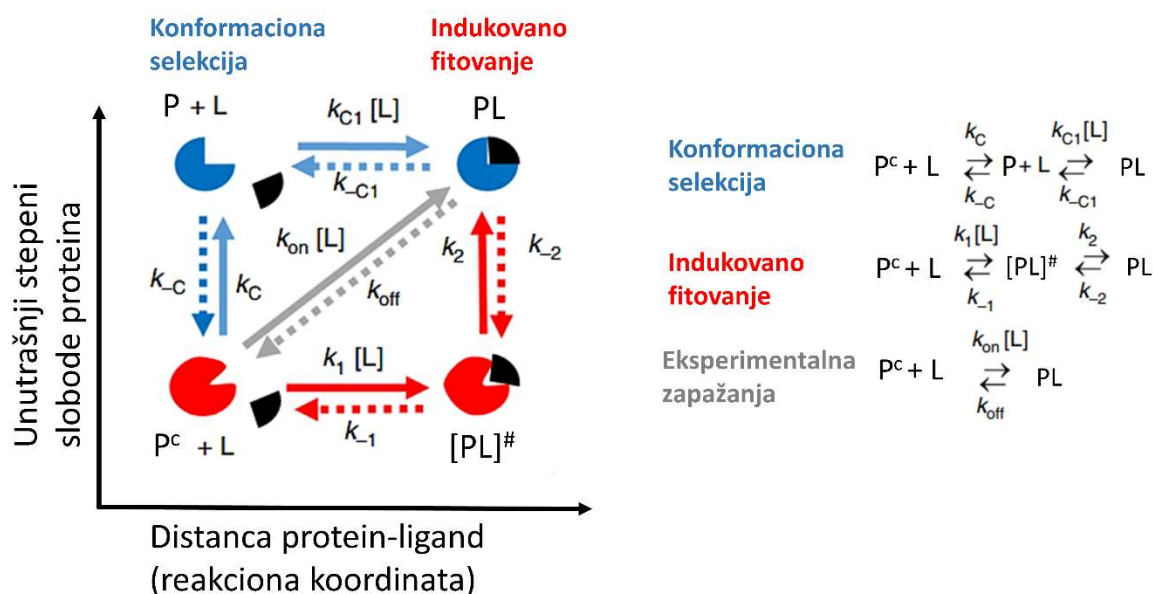
$$\Delta S = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t} \quad (17)$$

gde ΔS_{solv} predstavlja promenu entropije samog solventa koja je u najvećoj meri posledica izmene molekula rastvarača sa površine proteina usled vezivanja liganda. Doprimos ΔS_{solv} se uglavnom smatra povoljnim usled visoke pozitivne vrednosti, odnosno porasta entropije. ΔS_{conf} predstavlja promenu konformacione entropije koja je u vezi sa promenama u broju konformacionih stepeni slobode proteina i liganda nakon interakcije. Doprimos ΔS_{conf} se razlikuje od slučaja do slučaja jer se stepeni konformacione slobode liganda ili proteina mogu povećati, ili smanjiti usled interakcije. $\Delta S_{r/t}$ predstavlja gubitak translacionih i rotacionih stepeni slobode liganda i proteina nakon interakcije i generisanja kompleksa. Doprimos $\Delta S_{r/t}$ se često tumači kao nepovoljan usled smanjenja broja čestica u sistemu nakon formiranja kompleksa. U najvećem broju slučajeva, da bi do interakcije došlo, nepovoljna promena u $\Delta S_{r/t}$ entropiji se mora nadomestiti ili kroz veliki porast entropije solventa ΔS_{solv} , ili kroz veliki entalpijski doprimos nekovalentnih interakcija^{3,48}.

Afinitet liganda za određeni protein je proizvod balansa entropijskih i entalpijskih doprimosa ukupnoj energiji vezivanja. Entalpijski doprimos se može tumačiti kroz poznavanje 3D-strukture kompleksa i prirode intermolekulskih nekovalentnih interakcija. Prilikom dizajna novih liganada poboljšanog afiniteta ka ciljnom mestu, entalpijski doprimos se može modulirati ciljanim uvođenjem novih supstituenata u strukturu liganda kojima bi se uspostavile dodatne nekovalentne interakcija sa ciljnim mestom dejstva. Entropijski doprimos je značajno teže tumačiti, a čitav proces dizajna novih liganada se dodatno komplikuje usled fenomena poznatog kao entalpijska-entropijska kompenzacija. Entalpijska-entropijska kompenzacija se odnosi na promene u termodinamičkim veličinama entalpije i entropije tokom procesa vezivanja dva molekula. Entalpijska-entropijska kompenzacija se javlja kada promene u entalpiji i entropiji suprotstavljaju jedna drugoj, tako da ukupna promena slobodne energije ostaje relativno konstantna^{3,48,50}.

Postoje tri glavna konceptualna modela kojima se opisuje protein-ligand interakcija: model ključa i brave, model indukovano gfitovanja i model konformacione selekcije (Slika 1.6)⁵⁰:

- Osnovna pretpostavka modela ključa i brave je rigidnost vezivnog mesta i topološka komplementarnost obliku liganda. Ligand se uklapa u vezivno mesto poput ključa koji se uklapa u bravu. Ovaj model pretpostavlja da je protein statičan i da ne menja oblik nakon vezanja liganda što je u više navrata opovrgnuto eksperimentalnim studijama koje su ukazale na promene u konformacionom stanju proteina pre i nakon interakcije sa ligandima, zbog čega se danas smatra prevaziđenim^{3,51};
- U modelu indukovano gfitovanja, vezivno mesto na proteinu se smatra fleksibilnim i može se prilagoditi obliku liganda tokom interakcije. Ligand indukuje konformacionu promenu u proteinu, dovodeći do boljeg uklapanja između dva molekula. Ovaj model pretpostavlja da su i ligand i protein dinamični i mogu menjati konformaciono stanje. Ipak, prema ovom modelu protein ne podleže velikim konformacionim promenama što ograničava primenu modela na grupu relativno rigidnih proteina, ili određene faze tokom same interakcije^{3,51};
- Model konformacione selekcije predlaže da protein postoji u više konformacionih stanja, dok se ligand selektivno vezuje za specifično konformaciono stanje koje je povoljno za vezivanje. Drugim rečima, ligand stabilizuje protein u odabranom konformacionom stanju, dovodeći do čvršćeg vezivanja. Ovaj model proizilazi iz teorije polja slobodne energije kojom se objašnjava istovremeno postojanje različitih ansambala konfiguracija termodinamičkog sistema koji se međusobno nalaze u ravnoteži i čija verovatnoća pronalazjenja u faznom prostoru direktno korespondira sa razlikama u slobodnim energijama između mogućih stanja. Pojednostavljeno, protein se može određenom verovatnoćom naći u konformacionom stanju koje odgovara vezivanju liganda, iako ligand nije prisutan. Uvođenjem liganda u sistem, dolazi do stabilizacije proteina u konformaciji koja odgovara vezivnom stanju^{3,51}.



Slika 1.6. Šematski prikaz mehanizama protein (P)-ligand (L) interakcije – konformacione selekcije (plava putanja) i indukovano gfitovanja (crvena putanja). Siva putanja predstavlja eksperimentalna zapažanja. $k_{C/-C}$ su konstante brzine intrizičkih konformacionih promena proteina; $k_{2/-2}$ su konstante brzine konformacionih promena indukovanih procesom vezivanja/disocijacije liganda; $k_{1/-1}$ i $k_{C1/-C1}$ su konstante brzine formiranja protein-ligand kompleksa, dok su k_{on} i k_{off} eksperimentalno utvrđene vrednosti vezivanja, odnosno disocijacije liganda. Prilagođeno na osnovu reference⁵⁰.

Treba istaći da se indukovano fitovanje i konformaciona selekcija međusobno ne isključuju i da se čak u najvećem broju slučajeva sreće kombinacija ova dva mehanizma (Slika 1.6). Poznavanje ovih mehanizama, makar i načelno, može poslužiti u planiranju strategija računarskog dizajna novih liganada. Računarskim metodama za izučavanje konformacione dinamike moguće je dobiti detaljan uvid u mehanizam protein-ligand interakcije na atomskoj rezoluciji, uključujući i visokoenergetska prelazna stanja koja nije moguće izučavati eksperimentalnim metodama⁵².

1.3.1.b. Molekulski doking

U okviru metoda racionalnog dizajna koje se oslanjaju na poznavanje strukture ciljnog biomakromolekula, molekulski doking se izdvaja kao jedna od najšire korišćenih metoda. Molekulski doking je računarska metoda koja se primenjuje u cilju modelovanja 3D-strukture receptor-ligand kompleksa (gde receptor može predstavljati protein, RNK, ili DNK molekul), odnosno predviđanja najverovatnije poze (orijentacije i konformacije) liganda unutar svog vezivnog mesta na ciljnom molekulu.

Protokol molekulskog dokinga se sastoji iz dva procesa: generisanje mogućih poza pomoću algoritama za pretragu i rangiranja (eng. *scoring*) dobijenih poza pomoću definisanih scoring funkcija⁵³.

Algoritam za pretragu generiše različite orijentacije i konformacije liganada u vezivnom mestu proteina uzimajući u obzir torzione, rotacione, translacione stepene slobode samog liganda. S obzirom da bi testiranje svih mogućih stepeni slobode liganda zahtevalo previše vremena i računarske snage, algoritmi za pretragu koriste strategije stohastičke, ili sistematske pretrage kako bi efikasnije došli do optimalnog rešenja, odnosno stanja energetskega minimuma. Sistematska pretraga podrazumeva uzorkovanje vezivnih poza liganda postepenom promenom konformacionih parametara liganda u predefinisanim intervalima. Dve najzastupljenije varijante sistematskog algoritma za pretragu predstavljaju fragmentni algoritam i algoritam intenzivne pretrage. Fragmentni algoritam funkcioniše po principu postepenog građenja punog molekula od fragmenata u vezivnom mestu, dok algoritam intenzivne pretrage ligand posmatra kao celinu. Glavna mana ovih algoritama proističe iz sistematskog načina promene konformacionih parametara liganda zbog čega ligand može ostati zarobljen u lokalnom energetskega minimumu, odnosno izvan globalnog energetskega minimuma⁵⁴.

Stohastička pretraga podrazumeva uzorkovanje vezivnih poza liganda primenom nasumičnih modifikacija konformacionih parametara. Na ovaj način se značajno umanjuje problem poze zarobljene u lokalnim energetskega minimumima. Dva najzastupljenija tipa ovog algoritma predstavljaju Monte Carlo i genetski algoritam. U okviru Monte Carlo algoritma, nasumične promene u konformacionim parametrima liganda se evaluiraju primenom energetskega skora i ukoliko se razlika u energijama polazne i modifikovane poze uklapaju u kriterijum definisan *Boltzmann*-ovom distribucijom, modifikovana poza se prihvata⁵⁵. Sa druge strane, genetski algoritmi generišu najverovatnije poze liganda koristeći principe prirodne selekcije. Konformacioni parametri liganda se u okviru ovog algoritma posmatraju kao lista gena koji grade jedan hromozom, odnosno ligand. Mutacijom gena unutar polaznog hromozoma (liganda), generiše se populacija hromozoma rezultujući u novim pozama liganda koje se evaluiraju primenom scoring funkcije. Proces se ponavlja za sve hromozome koji imaju zadovoljavajuću vrednost scoring funkcije do iznalaženja optimalnog rešenja⁵⁶.

Sve poze koje generiše algoritam za pretragu podležu proceni primenom scoring funkcije. Generisani skorovi predstavljaju procenu afiniteta liganda u odgovarajućoj pozi za aktivno mesto. Postoje tri glavne grupe scoring funkcija: empirijske scoring funkcije, funkcije zasnovane na poljima sile i scoring funkcije zasnovane na znanju. Empirijske scoring funkcije razlažu skor u više energetskega komponenti interakcije (jonska veza, vodonična veza, hidrofobne interakcije itd.) kojima se dodaju odgovarajući koeficijenti izvedeni iz procesa treninga funkcije na setu jedinjenja poznatog afiniteta. Kvalitet ovih funkcija je u direktnoj vezi sa kvalitetom seta podataka

upotrebljenim za generisanje modela. Skoring funkcije zasnovane na molekulsko-mehaničkim poljima sile se ne oslanjaju na eksperimentalna merenja afiniteta, već za proračun koriste prethodno opisana molekulsko-mehanička polja sile. Skoring funkcije zasnovane na znanju se oslanjaju na interakcione potencijale izvedene iz statističke obrade kristalografski razrešenih protein-ligand kompleksa. Glavna pretpostavka ove grupe skoring funkcija jeste da će favorizovane interakcije imati iste ili slične geometrijske parametre u većini razrešenih protein-ligand kompleksa. Prednost ovog tipa skoring funkcije je ta što može da modeluje i određen broj nestandardnih interakcija⁵⁷.

Uopšteno govoreći, glavno ograničenje molekuskog dokinga leži u činjenici da sama metoda ne razmatra konformacionu dinamiku proteina, već se proteini posmatraju kao statične (ili samo delimično fleksibilne) strukture, zbog čega i generisani skorovi sa sobom nose određenu grešku i moraju se tumačiti sa dozom opreza^{58,59}. Zbog navedenog ograničenja metode, izučavanje ključnih protein-ligand interakcija primenom molekuskog dokinga, mora se izvoditi sa velikim oprezom i opravdano je samo u slučajevima kada je moguće izvršiti validaciju dobijenih poza na osnovu poređenja sa strukturno sličnim ligandom za koji postoji kristalografski razrešena struktura. U ostalim slučajevima, rezultat molekuskog dokinga se mora validirati rigoroznijom metodom molekuskih simulacija^{58,59}.

Sa druge strane, metoda je veoma brza i može da obradi hiljade liganada u veoma kratkom veremenskom roku zbog čega je molekulski doking glavnu primenu pronašao u virtuelnom skriningu baziranom na 3D-strukturi ciljnog mesta (eng. *Structure-based virtual screening* – SBVS).⁶⁰ Dodatno, molekulski doking se izuzetno često koristi kao alat za pripremu protein-ligand kompleksa pri izučavanjima konformacione dinamike primenom neke od metoda simulacija.

1.3.1.c. Virtuelni skrining

Virtuelni skrining (VS), računarski analog visokopropusnom skriningu (rutinskoj eksperimentalnoj metodi za otkrivanje novih *hit* struktura u farmaceutskoj industriji), zbog svoje ekonomičnosti i efikasnosti postao je jedna od najznačajnijih metoda u akademiji i farmaceutskoj industriji. U potrazi za novim ligandima, adekvatno postavljen VS protokol može da postigne slične, ili čak i bolje rezultate u odnosu na visokopropusni skrining⁶¹. VS, za razliku od visokopropusnog skrininga, pretražuje baze virtuelnih jedinjenja zbog čega mu je dostupan i znatno širi hemijski prostor za pretragu. Uzimajući u obzir veliku efikasnost i relativno nisku cenu ove metode, VS je postao prekretnica u savremenom otkriću lekova koja je dozvolila manjim kompanijama i akademiji da uđu u trku sa farmaceutskim gigantima. Pored svoje osnovne primene, VS je pronašao primenu i kao filter za prioritizaciju jedinjenja pre izvođenja *in vitro* skrining metoda⁶².

Metode virtuelnog skrininga se mogu podeliti na metode bazirane na poznavanju strukture liganda (eng. *Ligand-based virtual screening* – LBVS) i metode zasnovane na poznavanju 3D-strukture ciljnog mesta (SBVS). Obe grupe metoda problem otkrića novih liganada svode na problem klasifikacije i prepoznavanja aktivnih jedinjenja unutar velikog broja inaktivnih jedinjenja. Jedan od glavnih nedostataka VS metoda predstavlja mogućnost generisanja velikog broja lažno pozitivnih rezultata, zbog čega se metode VS moraju adekvatno validirati. Dva osnovna načina validacije VS metoda predstavljaju retrospektivni i prospektivni skrining^{63,64}. Validacija retrospektivnim skriningom se svodi na proces validacije modela pomoću jedinjenja koja su bila nevidljiva tokom procesa generisanja modela. Najčešće se validacioni set podataka kreira da što verodostojnije oponaša realne uslove primene VS modela. U tom kontekstu, veliki broj neaktivnih jedinjenja poznatijih kao „mameci“ se spaja sa malim brojem aktivnih jedinjenja i prati se sposobnost modela da iz validacionog seta podataka preopterećenog neaktivnim jedinjenjima oporavi što veći broj aktivnih jedinjenja. Statistički parametri koji se prate u retrospektivnoj validaciji modela se svode na parametre kojima se procenjuje kvalitet klasifikacionih modela i biće detaljnije predstavljeni u poglavlju 1.3.2. Prospektivni skrining podrazumeva primenu modela u realnim uslovima, odnosno izdvajanje potencijalno aktivnih jedinjenja skriningom baze podataka fizički dostupnih jedinjenja. *In vitro* potvrdom aktivnosti izdvojenih jedinjenja kvalitet modela se finalno potvrđuje^{63,64}.

Prednost SBVS metode u odnosu na LBVS metodu leži u mogućnosti istovremenog predviđanja i 3D-strukture kompleksa čime se omogućava i dalje izučavanje mehanizama protein-ligand interakcije što je neophodan korak u *hit-to-lead* optimizacijama. Metode LBVS se ističu po tome što omogućavaju izvođenje VS u slučajevima kada 3D-struktura ciljnog mesta nije poznata. Dodatno, zbog manjeg broja proračuna koje je potrebno izvesti, LBVS metode se smatraju bržim u odnosu na SBVS. Metode LBVS će biti detaljnije obrađene u poglavlju 1.3.2.

Metode VS se uglavnom izvode pristupom filtera, gde se sa svakim nivoom filtera pooštavaju kriterijumi za prolazak jedinjenja u narednu fazu. Pored standardnih filtera za procenu *drug-like*, *lead-like* osobina i predviđanje ADMET profila, često se među finalne filtere SBVS protokola uvode i rigoroznije i računarski zahtevnije metode molekulskih simulacija kojima se dodatno potvrđuje predviđena aktivnost^{65,66}.

Kao što je već diskutovano, molekulski doking predstavlja jednu od osnovnih metoda u SBVS. Ipak, treba istaći da se pod okriljem SBVS nalaze i druge metode koje za svoj rad zahtevaju poznavanje 3D-strukture ciljnog mesta. Jedan od široko zastupljenih pristupa SBVS se oslanja na preklapanju fingerprintova izvedenih iz struktura aktivnog mesta i struktura liganada⁶².

Sva ograničenja vezana za molekulski doking, koja proističu iz pretpostavke statičnosti proteina, se transliraju i na SBVS metode. Implementacijom informacije o konformacionoj dinamici proteina moguće je poboljšati efikasnost SBVS protokola. Spregom molekulsko dinamičkih simulacija sa SBVS kroz implementaciju alternativnih konformacionih stanja teorijski je moguće proširiti i hemijski prostor *hit* jedinjenja⁶⁷⁻⁷⁰.

1.3.1.d. Primena molekulsko dinamičkih simulacija u SBDD

Kao što je već diskutovano u poglavlju 1.1, simulacije molekulske dinamike predstavljaju zlatni standard u izučavanju konformacionog ponašanja biomakromolekulskih sistema, njihovih termodinamičkih, kao i kinetičkih osobina. U okviru SBDD, najznačajnija primena simulacija MD se svodi na rasvetljavanje mehanizama protein-ligand interakcije, kao i u izučavanju termodinamike samog procesa. Primenom opsežnih MD simulacija, ili srodnih metoda kojima se prevazilazi problem ograničenog uzorkovanja, moguće je istražiti veliki broj različitih konformacionih stanja dostupnih za vezivanje liganada koja ostaju nevidljiva za eksperimentalne metode. Ova stanja, pored toga što daju uvid u detaljan mehanizam protein-ligand interakcije, takođe mogu poslužiti u dizajnu potpuno novih grupa liganada. Najveći doprinos ispitivanja konformacione dinamike u SBDD se očekuje kod fleksibilnih proteina kod kojih postoji više konformacionih stanja razdvojenih niskim energetskim barijerama u kojima ligand može da zauzme više različitih bioaktivnih konformacija⁶⁷⁻⁷⁰. Takođe, veliki doprinos se očekuje i kod proteina koji učestvuju u protein-protein interakcijama gde je površina na kojoj se dešava interakcija velika, fleksibilna i slabo definisana. Ovakvi proteini predstavljaju ujedno i veliki izazov za racionalni dizajn lekova i svaka dodatna informacija u vezi sa mehanizmom protein-ligand i protein-protein interakcije predstavlja veliki iskorak unapred.⁷¹⁻⁷³

Kraće MD simulacije se često primenjuju kao dodatna potvrda relevantnosti poze liganada predviđenih molekulskim dokingom, ili za dodatno korigovanje predviđenih poza. Kao teorijski rigoroznija metoda u odnosu na molekulski doking, kraće MD simulacije mogu da prepoznaju lažno pozitivne poze iz molekulskog dokinga, ili da otklone greške u predviđenoj pozi kroz relaksaciju kompleksa i fina prilagođavanja bočnih aminokiselinskih ostataka prema strukturi liganada⁵⁹.

Sa druge strane, MD simulacije mogu da se koriste i kao rigorozna metoda za procenu afiniteta liganada ka ciljnom mestu. Kao što je već diskutovano, skor generisan molekulskim dokingom ima vrlo ograničenu primenu u proceni afiniteta liganada (poglavlje 1.3.1.b), dok se metodama MD simulacija može dosta preciznije izračunati slobodna energija vezivanja, odnosno proceniti afinitet liganada ka ciljnom mestu⁷⁴.

Sa aspekta statističke termodinamike, proces interakcije liganda i receptora može se posmatrati kao reakcija u kojoj se ligand transformiše iz slobodnog stanja u kom interaguje sa molekulima rastvarača, do vezanog stanja gde primarno interaguje sa atomima receptora. Vezivno i nevezivno stanje liganda se ne mogu posmatrati kao izdvojena statična stanja već naprotiv, podrazumevaju kolekciju diskretnih mikrostanja sa različitim konformacijama liganda i/ili proteina različite energije. Prema osnovnim teorijskim promatranjima statističke termodinamike, skup ovakvih mikrostanja se može okarakterisati funkcijom gustine stanja. Gustina stanja je centralna funkcija statističke termodinamike kojom se uspostavlja ključna veza između mikroskopskog stanja sistema (konfiguracije sistema npr.) i njegovih makroskopskih karakteristika (temperatura, pritisak, entalpija, interna energija itd.)⁷⁵.

U analitičkoj formulaciji slobodne energije u izotermalno-izobarskom ansamblu (NPT – najčešći termodinamički ansambl tokom simulacija), slobodna energija sistema u jednom stanju se može izraziti jednačinom 18:

$$Q(N, P, T) = \frac{1}{h^{3N} N!} \int \dots \int e^{-\beta(H(p_1 \dots p_N, q_1 \dots q_N) + PV)} dV dp_1 \dots dp_N, dq_1 \dots dq_N \quad (18)$$

gde Q predstavlja particionu funkciju sistema, h Plank-ovu konstantu, dok je $H(p, q)$ Hamiltonian sistema u kom p predstavlja momentum, a q koordinate čestice⁷⁵.

Usled potrebe za integracijom svih stepeni slobode sistema definisanih Hamiltonijanom, računanje apsolutne slobodne energije se izvodi samo za manje sisteme. Za veće sisteme poput proteina, simulacije u realnom vremenu trajanja nisu u stanju da posete sve delove faznog prostora ispitivanog sistema, zbog čega se primenjuje princip računanja relativne promene slobodne energije između dva stanja sistema. Imajući u vidu da je slobodna energija funkcija stanja sistema i zavisi samo od početnog i krajnjeg stanja, računanjem razlike slobodne energije između dva krajnja stanja potire se potreba za posjećivanjem svih delova faznog prostora. Drugim rečima, oduzimanjem apsolutnih energija dva stanja sistema, dolazi do poništavanja greške proistekle iz nedovoljnog uzorkovanja⁷⁶.

Jedna od osnovnih zakonitosti statističke termodinamike govori o direktnoj vezi funkcije verovatnoće stanja ($p(x)$) sa slobodnom energijom, zbog čega se ova relacija (jednačina 19) primenjuje u proceni razlike slobodnih energija dva sistema. Ova relacija predstavlja i direktnu vezu na osnovu koje je moguće iz simulacija izračunati slobodnu energiju. Naime, iz konvergiranih simulacija sistema moguće je izračunati verovatnoću nalaženja sistema u odgovarajućim stanjima i pomoću relacije (20) dalje promenu u slobodnoj energiji pri prelasku sistema između stanja:

$$p(x) \propto e^{-G(x)/k_b T} \quad (19)$$

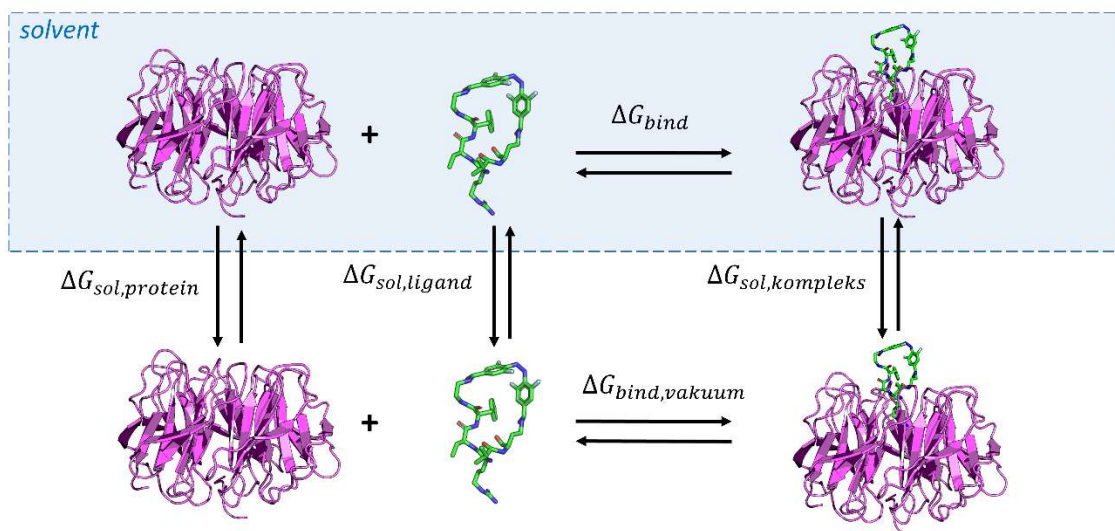
$$\frac{p(A)}{p(B)} = e^{-(G(A)-G(B))/k_b T} = e^{-\Delta G/k_b T} \quad (20)$$

gde je k_b Boltzmann-ova konstanta⁷⁵.

Dve najrigoroznije metode koje se oslanjaju na ovu relaciju su perturbacija slobodne energije (FEP) i termodinamička integracija (TI). Obe metode zahtevaju vrlo opsežne simulacije krajnjih stanja u termodinamičkom ciklusu i u cilju smanjenja greške uzorkovanja krajnjih stanja koriste tzv. alhemijsku, odnosno nefizičku, transformaciju sistema (npr. postepeni nestanak liganda iz vezivnog mesta, ili nestanak/pojavljivanje pojedinih atoma i veza u molekulu). Ove metode se još uvek nalaze u fazama validacije i njihova primena na konformaciono fleksibilnije sisteme je još uvek u fazi standardizacije zbog čega neće biti dalje obrađivane u sklopu ove teze⁷⁴.

Druga grupa metoda nazvanih metode krajnjeg stanja, koriste odgovarajuće aproksimacije radi uštede u računarskom vremenu, zbog čega su našle i širu primenu u strategijama SBDD. Među

najčešće korišćenim metodama krajnjeg stanja nalazi se i metoda molekulsko-mehaničke *Poisson–Boltzmann* površine (MM/PBSA). U okviru ove metode, molekuli solventa se posmatraju implicitno što značajno redukuje računarske resurse neophodne za proračune slobodne energije. Metode krajnjeg stanja se zasnivaju na termodinamičkom ciklusu prikazanom na Slici 1.7 i nose naziv po izvođenju iz klasičnih MD simulacija protein-ligand kompleksa. Iz jedne klasične MD simulacije protein-ligand kompleksa, ekstrakcijom trajektorije liganda i primenom implicitnog solventa, računa se slobodna energija solvatacije čime se eliminiše potrebe za bilo kakvim dodatnim simulacijama ponašanja samog liganda, ili *apo* proteina u solventu^{74,77}.



Slika 1.7. Termodinamički ciklus MM/PBSA metode. Plavim poljem je istaknuto polje solventa, dok polje vakuuma nije naznačeno bojom. $\Delta G_{sol,protein}$, $\Delta G_{sol,ligand}$, $\Delta G_{sol,kompleks}$ predstavljaju promene u slobodnoj energiji solvatacije (ΔG_{sol}) za protein, ligand i kompleks, redom (jednačina 23). $\Delta G_{bind,vakuum}$ predstavlja promenu u slobodnoj energiji vezivanja liganda u vakuumu i izražava se kroz kombinaciju doprinosa promene u molekulsko-mehaničkoj energiji (ΔE_{MM}) i entropijskog doprinosa ($T\Delta S$) (jednačina 21).

Slobodna energija vezivanja liganda (ΔG_{bind}) se u okviru prikazanog termodinamičkog ciklusa može razložiti na entalpijski i entropijski doprinos, pri čemu se entalpijski doprinos dalje može razložiti na promenu molekulsko-mehaničke energiju u vakuumu (ΔE_{MM}) i promenu slobodnu energiju solvatacije (ΔG_{sol}) nakon vezivanja liganda (jednačina 21).

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (21)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (22)$$

$$\Delta G_{sol} = \Delta G_{PB} + \Delta G_{SA} \quad (23)$$

$$\Delta G_{SA} = \gamma \cdot SASA + b \quad (24)$$

Molekulsko-mehanički doprinos je moguće izračunati primenom klasičnih polja sile, dok se promena u slobodnoj energiji solvatacije nakon vezivanja liganda (ΔG_{sol}) dalje računa na osnovu doprinosa elektrostatičkih efekata ΔG_{PB} (izračunatih primenom Poisson–Boltzmann implicitnog modela solventa) i doprinosa nepolarnih efekata ΔG_{SA} (izračunatih primenom računanja površina dostupnih solventu – SASA) (jednačine 22-24). Entropijski doprinos ($T\Delta S$) se najčešće zanemaruje zbog velikog računarskog utroška usled čega se MM/PBSA metoda može koristiti prvenstveno za kalkulaciju relativne slobodne energije vezivanja.^{74,77}

1.3.2. Dizajn lekova baziran na poznavanju strukture liganda

Dizajn baziran na poznavanju strukture liganda (LBDD) obuhvata skup metoda i tehnika koje se oslanjaju isključivo na informacije o strukturi i aktivnosti prethodno identifikovanih liganda. Dva fundamentalna pristupa u LBDD su: 1) dizajn, ili selekcija novih liganada na osnovu hemijske sličnosti sa jednim, ili više poznatih liganada; 2) dizajn, ili selekcija novih liganada uspostavljanjem modela kvantitativnog odnosa strukture i aktivnosti (QSAR) pomoću serije poznatih liganada. Prvi pristup ima veliku primenu u slučaju izostanka dovoljno velikog set podataka za formiranje QSAR modela. Pristup dizajnu, ili selekciji liganda na osnovu hemijske sličnosti vrši se procenom sličnosti jedinjenja sa jednim, ili više poznatih liganada na osnovu poređenja hemijskih struktura iskazanih odgovarajućim bročanim, odnosno vektorskim zapisom (poglavlje 1.3.2.a). Sličnost jedinjenja se dalje iskazuje odgovarajućim skorom koji je proporcionalan sa nivoom preklapanja vektorskih zapisa. Izdvojeni slučaj poređenja na osnovu sličnosti predstavlja pristup farmakofornog preklapanja gde se u obzir uzima i konformaciono ponašanje poznatog/ih liganda/ada na osnovu kog se definiše farmakofora, odnosno set hemijskih i strukturnih osobina relevantnih za interakciju sa ciljnim mestom. Jedna od osnovnih primena ovog pristupa jeste virtuelni skrining zasnovan na strukturi liganda (eng. *Ligand-based Virtual Screening* – LBVS)⁷⁸.

QSAR metodom se uspostavlja kvantitativna veza strukture i aktivnosti koja je od velikog značaja u dizajnu analoga poznatih jedinjenja. Osnovno ograničenje QSAR metode predstavlja neophodnost postojanja kvalitetnog seta podataka za generisanje QSAR modela. Usled napretka algoritama mašinskog učenja i sve većeg broja javno dostupnih setova podataka za trening modela, QSAR danas nalazi veliku primenu i u LBVS. Ova metoda, zbog svog kvantitativnog uvida i preciznosti, predstavlja jednu od najcenjenijih metoda racionalnog dizajna zbog čega će biti detaljnije obrađena u poglavlju 1.3.2.a⁷⁹⁻⁸¹.

Dok metode SBDD nude atomistički uvid u interakciju proteina i liganda, kompletan uvid u prirodu interakcije ponekad može izostati usled nedovlnog uzorkovanja konformacionih stanja sistema. Ovo je naročito značajno kod proteina veće inherentne fleksibilnosti. Sa druge strane, LBDD metode nude elegantna rešenja koja, oslanjajući se na matematičke, ili modele mašinskog učenja, mogu razotkriti skrivenu vezu između strukture liganda i njegove aktivnost. Na ovaj način se svi relevantni fenomeni protein-ligand interakcije objedinjuju pod okrilje istog modela, što uprkos limitiranoj interpretabilnosti, pruža mogućnost prevazilaženja greške prouzrokovane ograničenjima SBDD pristupa. Brojne studije su zapravo potvrdile komplementarnost LBDD i SBDD pristupa, zbog čega se u savremenim projektima dizajna novih lekova sve češće koristi kombinovani LBDD/SBDD pristup⁸². Sa porastom dostupnih baza podataka, porastao je i broj dostupnih setova za QSAR modelovanje što otvara veliku mogućnost za ubrzavanje dizajna novih lekova primenom kombinovanog pristupa.

1.3.2.a. Kvantitativni odnos strukture i aktivnosti

Razvijena pre više od 50 godina od strane Hanscha i Fujite, metoda ispitivanja kvantitativnog odnosa strukture i aktivnosti (QSAR) predstavlja metodu dizajna lekova zasnovanu na poznavanju strukture liganada.⁸³ QSAR modelovanje danas podrazumeva uspostavljanje računarskih, ili matematičkih modela kojima se pronalazi statistički značajna veza između hemijske strukture i kontinuirane (pIC_{50} , pEC_{50} , K_i , itd.) ili kategoričke (aktivna, neaktivna, toksična, netoksična, itd.) biološke osobine korišćenjem različitih metoda regresije i klasifikacije⁸⁴. QSAR tehnika predstavlja jednu od najuspešnijih strategija savremenog racionalnog dizajna lekova kojom se efikasno izbegavaju neuspesi u prekliničkom razvoju lekova zbog čega je i visoko cenjena u farmaceutskoj industriji^{44,79,84,85}.

Za izgradnju QSAR modela potrebni su visoko kvalitetni setovi podataka sastavljeni od hemijskih struktura jedinjenja i njihovih bioloških aktivnosti. Neposredno pre podele inicijalnog seta na trening set (set jedinjenja za generisanje modela) i test set (set jedinjenja za eksternu validaciju modela), neophodno je hemijsku strukturu jedinjenja prevesti u odgovarajući numerički (molekulski

deskriptor), ili vektorski izraz (fingerprint). Molekulski deskriptori predstavljaju brojčani izraz odgovarajućih fizičko-hemijskih i/ili strukturnih osobina jedinjenja, dok molekularni fingerprintovi predstavljaju binarni vektorski zapis odgovarajućih strukturnih karakteristika ispitivanog jedinjenja^{86,87}.

Odabir odgovarajućih molekularnih deskriptora i fingerprintova ključan je za izgradnju QSAR modela. Postoji nekoliko tipova molekularnih deskriptora, uključujući 1D, 2D, 3D i 4D deskriptore. 1D deskriptori, kao što su molekularna masa i broj donora i akceptora vodonične veze, računaju se iz molekularne formule jedinjenja i predstavljaju najjednostavniji tip deskriptora. 2D deskriptori, kao što su topološki indeksi i indeksi povezanosti, računaju se iz 2D topologije molekula, odnosno iz molekularnog grafa. 3D deskriptori, kao što su molekularni oblik i površina elektrostatičkog potencijala, za svoje proračune zahtevaju poznavanje konformacije molekula i zajedno sa 4D deskriptorima predstavljaju računarski najzahtevniji tip deskriptora. 4D deskriptori su dobili naziv po četvrtoj dimenziji koja najčešće oslikava višestruka konformaciona stanja samog molekula⁸⁸.

Fingerprintovi se uglavnom izvode iz 2D, ili 3D strukture molekula. Iako 3D fingerprintovi predstavljaju noviju grupu deskriptora koja sa sobom nosi veći nivo informacija, studije poređenja su pokazale da su 2D i 3D deskriptori podjednako dobar izbor pri modelovanju odnosa strukture i aktivnosti, pri čemu 2D deskriptori predstavljaju racionalni izbor pri radu sa velikim setovima podataka zbog veće brzine proračuna. MACCS i ECFP pripadaju grupi najpouzdanijih i najčešće korišćenih 2D fingerprintova⁸⁹.

- MACCS (eng. *Molecular ACCess System* – MACCS) su vrsta binarnih fingerprintova koji kodiraju prisustvo ili odsustvo skupa predefinisanih strukturnih obeležja u molekulu, stvarajući binarni niz fiksne dužine od 166 bita. Obeležja kodirana u MACCS fingerprintu uključuju funkcionalne grupe, prstenaste sisteme i druge uobičajene strukturne motive, kao što su heteroatomi azota i kiseonika. Prisustvo ili odsustvo svakog obeležja kodira se kao 1 ili 0, što rezultira binarnim fingerprintom koji predstavlja molekularnu strukturu na pojednostavljen i standardizovan način. Jedna od glavnih prednosti MACCS fingerprintova je njihova jednostavnost i lakoća korišćenja. Mogu se brzo izračunati za veliki broj molekula, čineći ih idealnim za virtuelni skrining i druge visokopropusne primene. Pored toga, njihovo kodiranje fiksne dužine čini ih pogodnim za efikasno poređenje i pretragu korišćenjem binarnih mera sličnosti, kao što je Tanimoto koeficijent sličnosti. Uzimajući u obzir visok diverzitet hemijskih struktura, fiksni skup unapred definisanih karakteristika koji se enkodiraju u MACCS fingerprint ne može da obuhvati sve relevantne strukturne informacije što predstavlja i glavno ograničenje ovog fingerprinta⁹⁰.
- ECFP (eng. *Extended Connectivity Fingerprint* – ECFP) je popularan tip binarnog cirkularnog fingerprinta koji beleži okruženje svakog atoma u predefinisanim radijusima. ECFP predstavljaju jedinstveni i kompaktan digitalni prikaz hemijske strukture molekula. ECFP4 se generišu iterativnim proširivanjem susedstva oko posmatranog atoma u molekulu do definisanog radijusa, uzimajući u obzir vrste hemijskih veza i prisutnih substrukture. Kroz svaku iteraciju definiše se određena cirkularna substruktura. Rezultat je niz bitova (najčešće 1024, ili 2048) koji kodiraju prisustvo ili odsustvo specifičnih podstrukture unutar molekula, pri čemu svaki bit predstavlja različiti fragment ili podstrukturu. ECFP fingerprintovi su široko korišćeni u različitim aplikacijama hemioinformatike, uključujući pretragu sličnosti, klasterovanje i modelovanje zasnovano na mašinskom učenju. Jedna od ključnih prednosti ECFP fingerprinta je njihova sposobnost da efikasno predstavljaju veliki broj molekula u kompaktnom formatu, što ih čini pogodnim za obradu velikih baza podataka. Za razliku od MACCS fingerprinta, ECFP nije ograničen predefinisanim setom substrukture i fragmenata zbog čega se može nesmetano koristiti u radu sa strukturama koje poseduju nove/netradicionalne funkcionalne grupe ili fragmente⁹¹.

U svojim počecima, QSAR modelovanje je bilo ograničeno na male serije strukturno sličnih jedinjenja i jednostavne metode regresije. Ovakvi QSAR modeli su se prevashodno koristili u svrhe dizajna analoga ispitivane serije jedinjenja. U poslednjih par decenija, koncept QSAR-a je pretrpeo mnoge promene usled razvoja naprednih algoritama mašinskog učenja i izgradnje velikih baza podataka, zbog čega je i primena ove tehnike danas značajno proširena^{80,81,84,92}. Retrospektivnim analizama pokazano je da se sa povećanjem raznovrsnosti i broja jedinjenja u trening setu modela može značajno povećavaju i tačnost, kao i primenljivost predviđanja. QSAR modeli trenirani na velikim setovima podataka sačinjenim od hemijski raznorodnih grupa jedinjenja mogu imati i širi domen primenljivosti, odnosno širu pokrivenost hemijskog prostora. Savremeni algoritmi mašinskog učenja, koji mogu efikasno uspostaviti vezu strukture i aktivnosti iz kompleksnih i velikih setva podataka, igraju značajnu ulogu u formiranju QSAR modela.⁹³ Neki od najpopularnijih algoritama su predstavljeni u poglavlju 1.3.3. Zbog pretpostavljenih prednosti, globalni QSAR modeli (QSAR modeli trenirani na velikim setovima podataka) postaju sve popularniji u savremenom otkriću lekova i osim u dizajnu analoga, pronalaze sve veću primenu i u LBVS^{80,81,84,92}.

Jedna od kritičnih determinanti kvaliteta QSAR modela predstavlja kvalitet seta podataka upotrebljen za kreiranje modela. U idealnom slučaju, biološke aktivnosti unutar seta podataka, pored raspona aktivnosti koji je minimum 5 puta veći od eksperimentalne greške, bi trebalo da budu određene i pod istim uslovima⁹⁴. Na ovaj način se eliminiše varijabilnost unutar podataka prouzrokovana različitim uslovima bioloških eseja, koja u velikoj meri može narušiti kvalitet QSAR modela. U realnom slučaju, ovakvi setovi podataka za QSAR modelovanje postoje samo za relativno mali broj strukturno sličnih jedinjenja i rezervisani su za kreiranje QSAR modela u studijama *lead* optimizacije. QSAR modeli za primenu u LBVS moraju da budu robusni i sa širokim domenom primenljivosti što se ne može postići treniranjem modela na malom setu strukturno sličnih jedinjenja. Sa druge strane, veliki setovi podataka, pohranjeni u bazama podataka kao što je ChEMBL⁹⁵, se gotovo uvek sastoje od bioloških aktivnosti testiranih pod različitim uslovima zbog čega se kvalitet ovih modela mora potvrditi opsežnim validacijama⁸⁴.

Kvalitet svakog QSAR modela se, prema OECD smernicama, potvrđuje metodama interne i eksterne validacije. Parametri koji se prate tokom validacije zavise od vrsta modela (regresioni, ili klasifikacioni)⁹⁶. Regresioni modeli su osetljivi na varijabilnost u eksperimentalnim podacima zbog direktnog modelovanja vrednosti zavisne promenljive, odnosno biološke aktivnosti. Sa druge strane, klasifikacioni modeli, usled podele jedinjenja u klase aktivnih i neaktivnih jedinjenja, su značajno manje osetljivi na uticaj varijabilnost u eksperimentalnim podacima. Usled veće osetljivosti, za regresione modele do danas je razvijen i veći broj validacionih parametara koji sagledavaju kvalitet modela sa više različitih aspekata⁹⁷.

Interna validacija regresionih QSAR modela se izvodi praćenjem koeficijenta determinacije (R^2), i korelacionog koeficijenta ukrštene validacije (Q^2), korena srednje kvadratne greške (eng. *Root Mean Square Error* – RMSE) fitovanja trening seta ($RMSE_{int}$) i ukrštene validacije ($RMSE_{cv}$). Svi parametri interne validacije se računaju pomoću seta podataka upotrebljenog za formiranje modela (trening set). Dodatna potvrda validnosti modela može se izvršiti protokolom Y-randomizacije kojim se generiše veliki broj modela permutacijama zavisne varijable unutar seta podataka. Poređenjem permutovanih modela sa originalnim modelom utvrđuje se da li su predviđanja originalnog QSAR modela posledica stvarne povezanosti strukture i aktivnosti, ili samo slučajnosti.^{98,99}

Eksterna validacija modela je mnogo značajniji parametar kvaliteta modela. Izvodi se upotrebom test seta, odnosno seta podataka koji nije korišćen za generisanje modela. Iako OECD smernice predlažu primenu Q_{F1}^2 za eksternu validaciju modela, veći broj studija je utvrdio ograničenja ovog parametra zbog čega je preporučena primena i ostalih validacionih parametara poput: koeficijent determinacije test seta (R_{ext}^2), RMSE test seta ($RMSE_{ext}$), Q_{Fn}^2 metrika (Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2), r_m^2 metrika (r_m^2 , \bar{r}_m^2 , Δr_m^2) i korelacioni koeficijent podudarnosti (eng. concordance correlation coefficient (CCC)).^{100–103}

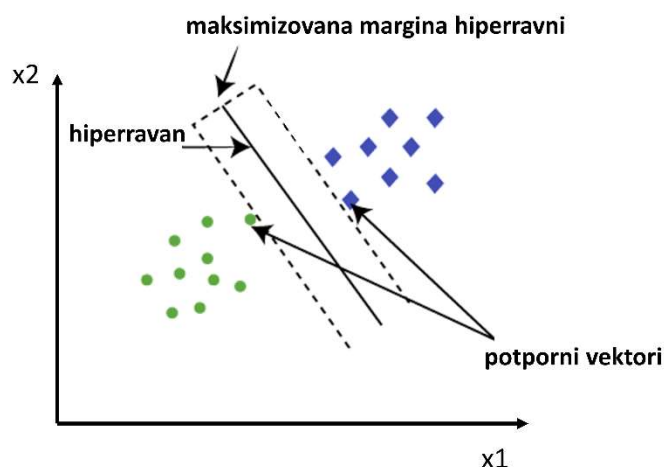
Klasifikacioni modeli se uglavnom evaluiraju koristeći metriku izvedenu iz konfuzionih matrica formiranih praćenjem broja pravih pozitivnih predviđanja (eng. *True Positive* – TP), pravih negativnih predviđanja (eng. *True Negative* – TN), lažno pozitivnih predviđanja (eng. *False Positive* – FP) i lažno negativnih predviđanja (eng. *False Negative* – FN). Najčešće korišćeni parametri za internu i eksternu validaciju modela su: balansirana tačnost (eng. *Balanced Accuracy* – BA), opoziv (eng. *Recall*), preciznost, F1-skor, *Matthews*-ovog korelacionog koeficijenta (eng. *Matthew's correlation coefficient* – MCC) i površina ispod ROC (eng. *Receiver Operating Characteristics* – ROC) krive (ROC_AUC)^{104,105}.

1.3.3. Algoritmi mašinskog učenja u otkriću lekova

Otkriće novih lekova u okviru inicijative za preciznu medicinu uveliko se oslanja na integraciju velikih skupova podataka u heminformatičke protokole⁴⁴. Era vođena velikim podacima u modernom otkrivanju lekova prepoznala je veštačku inteligenciju (eng. *Artificial Intelligence* – AI) kao jedan od najvažnijih alata koji može drastično smanjiti vreme i troškove otkrivanja lekova u pretkliničkim fazama^{79,80}. Eskalacija razvoja i upotrebe alata mašinskog učenja u modernom otkrivanju lekova uglavnom je omogućena dostupnošću velikih skupova podataka i demokratizacijom AI. Javne baze podataka farmakološki aktivnih jedinjenja, poput ChEMBL baze⁹⁵, sa rastućim brojem zapisa o biološkim aktivnostima omogućile su sveobuhvatan pristup modernom otkrivanju lekova implementacijom algoritama mašinskog učenja u modelovanju odnosa između strukture i aktivnosti, ali i strukturno zavisnom dizajnu lekova^{44,79,80}. U okviru ovog poglavlja predstavljeni su neki od najčešće korišćenih algoritama mašinskog učenja u okviru racionalnog dizajna lekova.

1.3.3.a. Algoritam potpornih vektora

Algoritam potpornih vektora (eng. *Supporting Vector Machines* – SVM) predstavlja često korišćenu klasu algoritama za nadgledano učenje koji se koriste za klasifikacionu i regresionu analizu. SVM je baziran na ideji pronalaženja hiperravni unutar N-dimenzionalnog prostora (gde N predstavlja broj nezavisnih varijabli) koja najbolje razdvaja različite klase u datom skupu podataka. Potporni vektori predstavljaju tačke unutar N-dimenzionalnog prostora, odnosno N-dimenzionalne vektore, pomoću kojih je definisana hiperravan, odnosno koji se nalaze najbliže definisanoj ravni (Slika 1.8). SVM algoritam teži da maksimizuje ovaj razmak, a istovremeno klasifikuje sve podatke na ispravan način (Slika 1.8). Osim maksimizovanja razmaka, SVM takođe uključuje regularizacioni član koji kažnjava model za donošenje prekomplikovanih odluka¹⁰⁶. Ovo pomaže u sprečavanju prevelikog prilagođavanja modela trening podacima i posledično loše performanse na eksternim podacima (eng. *overfitting*).



Slika 1.8. Princip algoritma potpornih vektora.

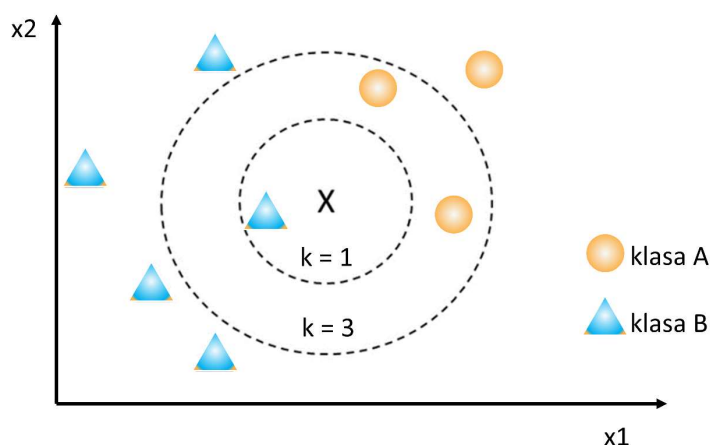
Jedna od ključnih prednosti SVM je sposobnost obrade visoko-dimenzionalnih podataka, kao što su molekularni deskriptori ili podaci o ekspresiji gena. To čini SVM posebno korisnim u aplikacijama poput otkrića lekova, gde su velike količine visoko-dimenzionalnih podataka česte. Pored toga, SVM je relativno računarski efikasan i može da obradi velike skupove podataka sa razumnim zahtevima za memorijom i vremenom¹⁰⁶⁻¹⁰⁸.

Postoji nekoliko varijanti SVM, zavisno od tzv. kernel funkcije koju algoritam koristi pri transformaciji ulaznih podataka. Neke od najpopularnijih varijanti uključuju linearni SVM, radijalnu baznu funkciju (eng. *radius basis function*) i sigmoidalni kernel. Linearni SVM dobro funkcioniše na setu linearno separabilnih podataka, dok ostale varijante mogu da obrade nelinearno separabilne podatke mapirajući ih u višedimenzionalnom prostoru. Multiklasni SVM proširuju osnovni SVM algoritam kako bi obradili skupove podataka sa više od dve klase¹⁰⁶⁻¹⁰⁸.

Ograničenja SVM uključuje osetljivost metode na izbor hiperparametara, kao što su parametar regularizacije i funkcija kernela, što može zahtevati pažljivo podešavanje da bi se postigle optimalne performanse¹⁰⁶⁻¹⁰⁸.

1.3.3.b. Algoritam k-najbližih suseda

Algoritam K-najbližih suseda (eng. *K-Nearest Neighbors* – KNN) je popularna tehnika mašinskog učenja koja se koristi za klasifikaciju i regresionu analizu. U svojoj osnovi, KNN je jednostavan i intuitivan algoritam koji se oslanja na ideju sličnosti između instanci u skupu podataka. Kada se novoj instanci dodeljuje klasa, KNN pretražuje skup podataka u potrazi za K-najbližih instanci (odnosno K najbližih suseda) i dodeljuje novu instancu većinskoj klasi među susedima (Slika 1.9). Broj suseda K je hiperparametar koji može da se podešava. Jedna od ključnih prednosti KNN-a je njegova jednostavnost i fleksibilnost¹⁰⁹⁻¹¹¹. Lako se implementira i može se primeniti na širok spektar problema, uključujući i klasifikaciju i regresione zadatke. KNN može da se nosi sa nelinearnim podacima koji prate kompleksne obrasce. Još jedna prednost KNN-a je sposobnost učenja iz lokalne strukture podataka, bez prilagođavanja globalnog modela. Drugim rečima, KNN model se neće promeniti ukoliko se novi podaci dodaju nakon treniranja modela, što nije slučaj sa ostalim metodama.



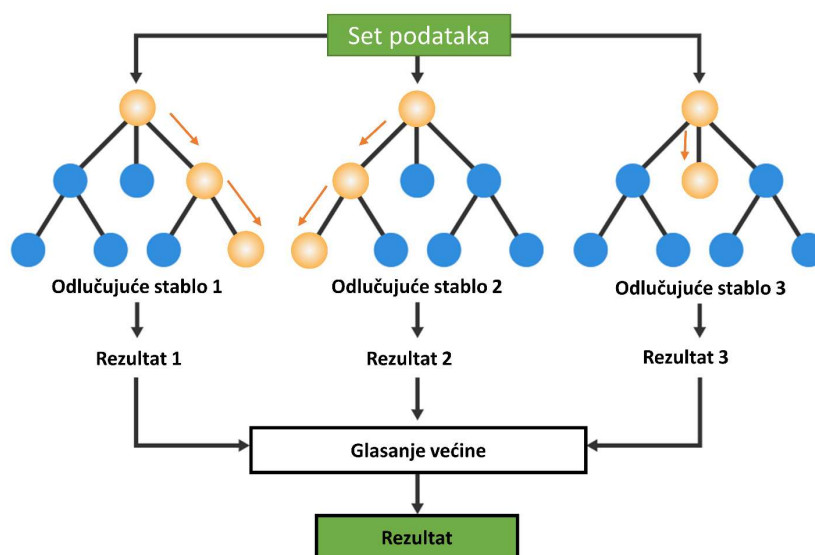
Slika 1.9. Princip rada KNN algoritma.

Jedno od ograničenja KNN podrazumeva računarsku složenost algoritma, koja raste sa veličinom i dimenzionalnošću skupa podataka. Ovo ga čini manje pogodnim za velike skupove podataka, posebno kada je broj nezavisnih varijabli visok. Osim toga, KNN može biti osetljiv na izbor metrike rastojanja, što može značajno uticati na njegove performanse¹⁰⁹⁻¹¹¹.

1.3.3.c. Algoritam slučajnih šuma

Slučajne šume (eng. *Random Forests* – RF) su algoritam mašinskog učenja koji se koristi za klasifikaciju, regresiju i druge prediktivne zadatke. Spada u tzv. metode mašinskog učenja iz ansambla odnosno u metode koje kombinuju ansambl od više modela (u ovom slučaju odlučujućih stabala) da bi se poboljšala tačnost i otpornost predviđanja¹¹².

U osnovi, RF algoritam je zasnovan na kombinovanju velikog broja pojedinačnih predviđanja odlučujućih stabala. Svako odlučujuće stablo se gradi tako što se nasumično odabere podskup podataka i model trenira za dati podskup. Konačno predviđanje se zatim vrši kombinovanjem predviđanja svih pojedinačnih stabala (Slika 1.10)¹¹².

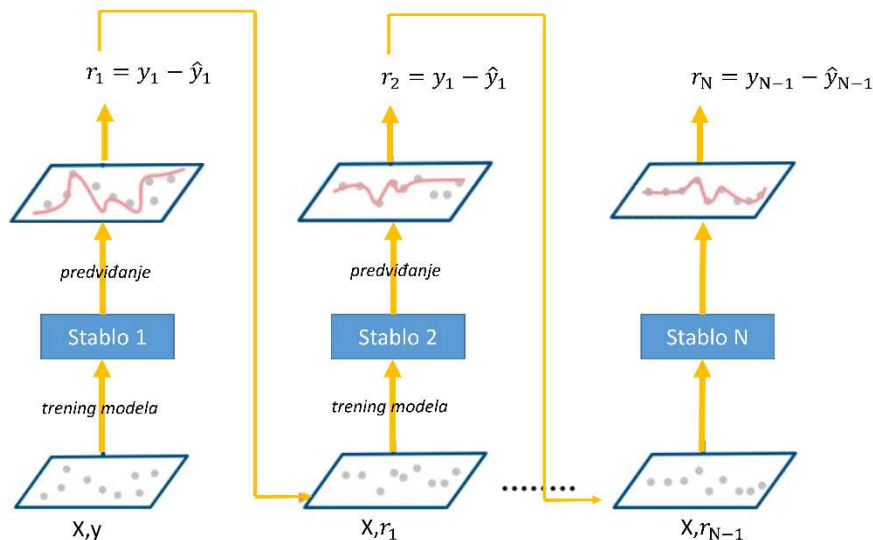


Slika 1.10. Princip rada RF algoritma.

Jedna od ključnih prednosti RF je sposobnost nošenja sa složenim, nelinearnim i visokodimenzionalnim podacima. Pored toga, algoritam pokazuje visoku robusnost odnosno relativnu neosetljivost na izbor hiperparametara, kao što su broj stabala, dubina stabala i veličina podskupa karakteristika. RF može da pruži procene relativne važnosti svake karakteristike/nezavisne promenjive u predviđanju, što može biti korisno za selekciju i interpretaciju nezavisnih promenjivih. Od svog uvođenja u QSAR modelovanje, RF je postao jedan od najuspešnijih algoritama i zlatnih standarda pri formiranju QSAR modela^{112,113}.

1.3.3.d. Algoritam ekstremnog gradijentnog pojačanja

Gradijentno pojačanje (eng. *Gradient boosting*) je još jedna vrsta ansambl metoda učenja koja kombinuje nekoliko slabijih modela, obično drveća odlučivanja, kako bi se stvorio jači model koji je sposoban da napravi preciznija predviđanja. Za razliku od algoritma slučajnih šuma koji finalni model dobija paralelnim treningom odlučujućih stabala, algoritmi gradijentnog pojačanja koriste sekvencionalni pristup (Slika 1.11)^{114,115}.



Slika 1.11. Princip rada algoritma gradijentnog pojačanja.

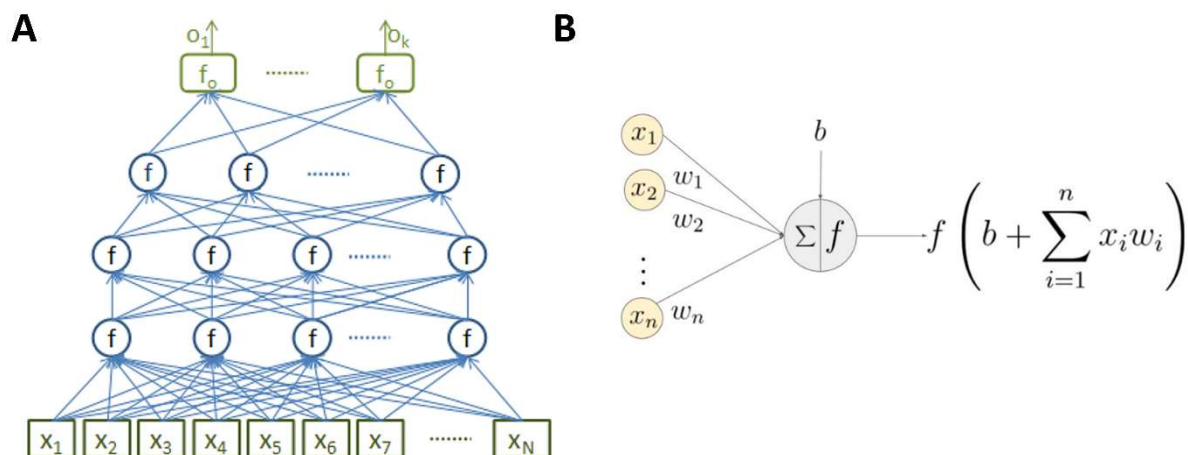
Osnovna ideja iza gradijentnog pojačanja je iterativno poboljšavanje modela dodavanjem novih stabala koja koriguju greške prethodnih stabala. Primenom metode gradijentnog spusta, algoritam minimizuje funkciju gubitka koja meri razliku između predviđenih i stvarnih vrednosti ciljane promenljive. U svakoj iteraciji, novo stablo se trenira na rezidualima, odnosno razlici između predviđenih i stvarnih vrednosti iz prethodne iteracije (Slika 1.11). Novo stablo se zatim dodaje u ansambl, a postupak se ponavlja sve dok se ne postigne željeni nivo tačnosti¹¹⁴⁻¹¹⁶.

Algoritmi gradijentnog pojačanja se nalaze među najučinkovitijim algoritmima. Postoji nekoliko varijacija gradijentnog pojačanja, uključujući *AdaBoost*, koji je originalni algoritam pojačanja, i *XGBoost* (eng. *eXtreme Gradient Boosting* – XGBoost), koji je naprednija verzija koja uključuje nekoliko naprednih funkcija, kao što su paralelno procesiranje i regularizacija kojom se sprečava pretreniranje modela (eng. *overfitting*)¹¹⁴⁻¹¹⁶.

1.3.3.e. Duboke neuronske mreže

Duboke neuronske mreže (eng. *Deep Neural Networks* – DNN) su vrsta veštačkih neuronskih mreža koje su poslednjih godina postale sve popularnije zbog svoje sposobnosti efikasnog učenja iz velikih setova podataka i rešavanja kompleksnih problema. Za razliku od klasičnog algoritma veštačkih neuronskih mreža, DNN se sastoje od velikog broja skrivenih slojeva povezanih neurona koji zajedno transformišu ulazne podatke u izlazna predviđanja. Ova razlika omogućuje DNN algoritmu da uči iz velikog broja ulaznih podataka i eliminiše potrebu za prethodnim filtriranjem nezavisnih promenljivih čime se sprečava odliv korisnih informacija iz samog seta podataka^{117,118}.

Zahvaljujući napretku u algoritmima optimizacije kao i hardverskim rešenjima, savremeni DNN algoritmi mogu automatski da uče na više nivoa apstrakcije, što im omogućava da opišu kompleksne odnose i obrasce unutar seta podataka. Najčešći tip DNN je neuronska mreža nepovratnog tipa (eng. *feedforward*), koja se sastoji od ulaznog sloja, jednog ili više skrivenih slojeva i izlaznog sloja i u kojoj protok informacija teče od ulaznog ka izlaznom neuronu. Svaki sloj sadrži skup neurona koji izvode transformaciju ulaznih podataka primenom odgovarajućeg težinskog i faktora pristrasnosti. Transformisani podaci se zatim prosleđuju odgovarajućoj aktivacionoj funkciji. Aktivacione funkcije predstavljaju jednostavne nelinearne matematičke operacije kojima se odlučuje o relevantnosti transformisanog podatka odnosno neurona u celokupnoj mreži (Slika 1.12). Izlaz svakog sloja se zatim unosi u sledeći sloj, što omogućava mreži da nauči sve složenije reprezentacije podataka. Koristeći izlazne informacije nakon prvog ciklusa treninga, računa se greška na osnovu koje se težinski faktori preračunavaju i ažuriraju unazad sloj po sloj računanjem gradijenta funkcije zavisnosti greške od težinskog faktora^{113,117,119,120}.



Slika 1.12. Princip funkcionisanja DNN algoritma. A – Prikaz arhitekture DNN mreže sa dva izlazna neurona; B – Prikaz jednog neurona u DNN modelu, gde x_1 - x_n predstavljaju ulazne podatke, w_1 - w_n težinske faktore, b – faktor pristrasnosti, f – aktivacionu funkciju. Prilagođeno na osnovu reference¹¹³.

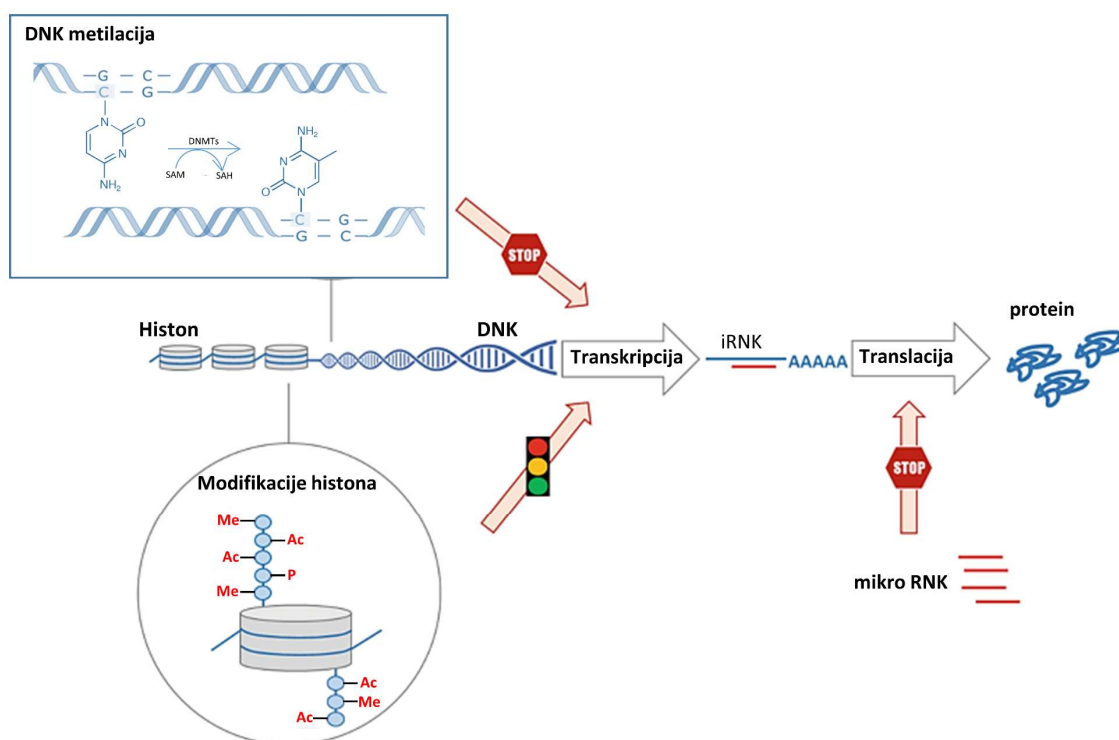
Jedan od izazova u treniranju DNN-a je prekomerno fitovanje (eng. *overfitting*), koji se javlja kada mreža postane previše složena i počne da prilagođava šum u trening podacima umesto osnovnih uzoraka. Da bi se sprečilo prekomerno fitovanje, razvijene su tehnike regularizacije, među kojima je najpopularnija tehnika isključivanja (eng. *dropout*) zasnovana na randomizovanom isključivanju neurana iz slojeva DNN-a tokom procesa treniranja mreže^{113,117,119,120}.

Osim neuronska mreža nepovratnog tipa, postoji nekoliko drugih tipova DNN koji su razvijeni, kao što su konvolutivne neuronske mreže (eng. *Convolutional neural network –CNN*) i rekurentne neuronske mreže (RNN). CNN su dizajnirani da obrađuju podatke u obliku mreže, kao što su slike, i koriste konvolutivne slojeve da nauče lokalne karakteristike. RNN, sa druge strane, su dizajnirani da procesiraju sekvencijalne podatke, poput teksta, ili tekstualnih zapisa hemijskih struktura, i koriste rekurentne veze da opišu temporalne zavisnosti¹¹⁷.

Sofisticiranost DNN algoritama danas igra veliku transformativnu ulogu u primeni veštačke inteligencije u farmaceutskoj industriji. DNN nalazi veliku primenu ne samo u QSAR analizi velikih baza podataka, već i u *de novo* dizajnu novih aktivnih supstanci, pretrazi novih ciljnih mesta i *off-target* studijama, kao i u sintetskom planiranju^{113,119,120}.

1.4. Epigenetska ciljna mesta u malignim tumorima

Pojam "epigenetika" je prvobitno uveo britanski biolog Conrad Hal Waddington u 1940-im godinama. Waddington je definisao epigenetiku kao: „Grana biologije koja proučava uzročne interakcije između gena i njihovih proizvoda, koji dovode do nastanka fenotipa”¹²¹. Waddington je koristio termin "epigenetska kontrola" da opiše način na koji se genetska informacija prenosi od gena do proteina, i kako se ti procesi mogu modifikovati tokom razvoja organizma. Ideja da se nasledne promene u genomu mogu javiti nezavisno od promena u DNK sekvenci bila je novina u to vreme i predstavljala je veliki iskorak u razumevanju genetike. Od tada, termin epigenetika je evoluirao i danas podrazumeva sve nasledne promene u funkciji gena koje se ne odnose na promene u DNK sekvenci, već su povezane sa strukturama van osnovnog genetskog materijala (grč. *epi-* na, preko, iznad) uključujući promene u DNK metilaciji, histonima i hromatinu, kao i regulaciju nekodirajućih RNK^{122,123}.



Slika 1.13. Prikaz osnovnih epigenetskih mehanizama. Prilagođeno na osnovu referenci ¹²⁴ i ¹²⁵.

Epigenetski mehanizmi su procesi koji regulišu način na koji se genetska informacija izražava u ćeliji, a koji se nasleđuju od ćelije do ćelije tokom razvoja i reprodukcije. Ovi mehanizmi ne menjaju DNK sekvencu gena, ali mogu modifikovati način na koji se geni ekspimiraju. Postoje tri glavna epigenetska mehanizma: DNK metilacija, modifikacija histona i regulacija nekodirajućih RNK (Slika 1.13)¹²⁶.

- **DNK metilacija:** DNK metilacija predstavlja jedan od najbolje izučenih epigenetskih mehanizama. Ovaj mehanizam uključuje kovalentno vezivanje metil grupe u položaju 5 pirimidonskog heterociklusa citozina. DNK metilacija se odigrava u delovima DNK poznatijim kao CpG-sekvence (citozin-gvanin dinukleotidna sekvenca) i primarno je katalizovana enzimima poznatijim kao DNK-metiltransferaze (DNMT). DNK metilacija, uz par izuzetaka, dovodi do represije ekspresije gena sprečavanjem pristupa transkripcionim faktorima ili rekonfiguracijom hromatina usled vezivanja proteina sa metil-vezujućim domenima. DNK demetilovanje predstavlja obrnuti proces kojim se reaktiviraju geni utišani DNK-metiltransferazama. DNK demetilovanje je katalizovano posebnom grupom metilcitozin dioksidogenaza poznatijih kao TET (eng. *ten-eleven translocation* – TET) metilcitozin dioksidogenaze koji oksiduju 5-metilcitozin do 5-hidroksimetilcitozina i dalje do 5-formilcitozina i 5-karboksilcitozina^{127–129}.
- **Modifikacija histona:** Unutar hromatina, delovi DNK su upakovani u visoko kondenzovanu strukturu sastavljenu od oktamera histonskih proteina (nukleozom). Histonski oktameri su sastavljeni od po dve kopije histona 2A, histona 2B, histona 3 i histona 4. C-terminalni i N-terminalni krajevi histonskih proteina podležu različitim posttranslacionim modifikacijama, uključujući acetilovanje, metilovanje, ADP ribozilaciju, fosforilaciju itd. kojima se direktno utiče na proces transkripcije DNK. Skup spomenutih modifikacija se, usled količine transkripcionih informacija koje u sebi nosi, često oslovljava i terminom „histonski kod“. Među spomenutim posttranslacionim modifikacijama, najdetaljnije su izučavane modifikacije lizinskih ostataka histona H3 i H4. Neutralizacijom pozitivnog naelektrisanja terminalnog azota lizina kroz acetilovanje, pospešuje se dekonverzovanje

hromatina i olakšava se pristup transkripcionim faktorima čime se ostvaruje kontrola na nivou transkripcije DNK. Glavni enzimi uključeni u regulaciju acetilacionog statusa hromatina predstavljaju histon-deacetilaze (HDAC) i histon-metiltransferaze (HAT). Sa druge strane, epigenetska regulacija transkripcije metilovanjem aminokiselinskih ostataka histona posredovana je različitim histon-metiltransferazama i demetilazama. Mehanizam epigenetske regulacije transkripcije DNK u slučaju metilovanja histona nije povezan sa direktnim efektom na stepen kondenzacije hromatina već sa prepoznavanjem od strane različitih histon-vezujućih proteina, tzv. „čitača“ epigenetskog koda^{126,130}.

- **Regulacija nekodirajućih RNK:** Nekodirajuće RNK (ncRNA) su RNK molekuli koji ne kodiraju proteine, ali igraju ključnu ulogu u regulaciji gena. Zauzimaju čak 70% genoma i dele se na duge ncRNK (lncRNK, > 200 nt) i kratke ncRNK (sncRNK, < 200 nt). Jedan od najpoznatijih podtipova sncRNK je mikro RNK (miRNK, 20-23 nt) koji učestvuje u regulaciji oko 60% gena u humanom genomu. miRNK svoju funkciju represora genske ekspresije ostvaruje kroz interakciju sa 3' UTR sekvencom odgovarajuće informacione RNK (iRNK) čime aktivira miRNK-indukovani utišavajući kompleks i time sprečava translaciju i podstiče degradaciju ciljane iRNK. Više od 50% gena miRNA je smešteno u blizini CpG-ostrvaca što njihovu ekspresiju čini podložnom epigenetskoj regulaciji¹³¹. Grupa lncRNA je prilično raznolika po strukturi i funkciji. Neke od funkcija lncRNA uključuju i regulaciju remodelovanja hromatina kroz interakciju sa različitim epigenetskim proteinima, poput intrakcije DNMT sa lncRNA koja direktno učestvuje u regulaciju metilacionog statusa¹³².

Na osnovu fiziološke uloge, proteini uključeni u navedene epigenetske mehanizme mogu se podeliti u tri grupe: pisači (eng. *writers*) (npr. DNK-metiltransferaze, histon-acetiltransferaze), brisači (eng. *erasers*) (npr. histon-deacetilaze) i čitači epigenetskog koda (eng. *readers*) (npr. BET (eng. *bromodomain and extraterminal* – BET) familija proteina)¹³³. Pisači, kao što i samo ime kaže, predstavljaju proteine koji dodaju epigenetske modifikacije na histone, ili DNK. Brisači uklanjaju navedene modifikacije, dok su čitači proteini koji prepoznaju navedene modifikacije i pomažu regrutovanju ostalih faktora koji služe kao transkripcioni koaktivatori, ili korepresori¹³⁴.

Svi ovi mehanizmi mogu biti pod uticajem spoljnih faktora, kao što su ishrana, toksini, stres i drugi stresori. Nedostatak ili prekomerna aktivnost epigenetskih mehanizama može dovesti do različitih bolesti, uključujući maligne tumore i druge poremećaje. U poslednjih 35 godina, epigenetske promene su prepoznate kao jedan od glavnih pokretača karcinogeneze i progresije malignih tumora¹³⁵. Inaktivacija tumor-supresorskih gena hipermetilacijom promotora u CpG ostrvcima je jedan od najbolje okarakterisanih mehanizama epigenetske regulacije velikog broja malignih tumora¹²⁷. Dodatno, disregulacija epigenetiskih mehanizama mutacijom i/ili aberantnom ekspresijom hromatin-modifikujućih enzima kao i proteina i nekodirajućih RNK molekula povezanih sa kontrolom strukture i dinamike hromatina takođe predstavlja važan aspekt epigenetske regulacije karcinoma. Uopšteno govoreći, epigenetska promena može dovesti do aktivacije onkogena, ili utišavanja tumor supresorskih gena, pomažući rast i preživaljavanje tumorske ćelije. Epigenetski mehanizmi takođe mogu da utiču i na rezistenciju tumorske ćelija na hemioterapiju¹²⁶.

Otkrićem centralne uloge koju epigenetska disregulacija igra u različitim vrstama karcinoma, nastalo je i veliko interesovanje za razvoj antitumorskih lekova koji interaguju sa epigenetskim mehanizmima. Trenutno se u kliničkoj praksi koristi devet antitumorskih lekova: inhibitori DNMT-a – azacitidin i decitabin; inhibitori izocitrat-dehidrogenaze 1/2 (IDH1/2) kojima se indirektno reguliše aktivnost TET-metilcitozin dioksigenaza – ivosidenib i enasidenib; inhibitor histon metiltransferaze EZH2 (eng. *Enhancer of zeste homolog 2* – EZH2) – tazetostat; inhibitori histon-deacetilaza (eng. *Histon deacetylase* – HDAC) – vorinostat, romidepsin, belinostat i panobinostat. Pored lekova odobrenih od strane Agencije za hranu i lekove Sjedinjenih Američkih Država (eng. *U.S. Food and Drug Administration* – FDA), veliki broj lekova koji ciljaju različita epigenetska ciljna mesta dospela su u kliničke studije, uključujući: inhibitor DOT1L histon-metiltransferaze

(eng. *DOT1 Like Histone Lysine Methyltransferase* – DOT1L) – pinometostat; menin-MLL (eng. *Mixed lineage leukemia protein* – MLL) inhibitori VTP-50469 i SNDX-5613; inhibitori LSD1 histon-demetilaze (eng. *Lysine-specific histone demethylase 1* – LSD1) – ORY-1001 i GSK2879552; inhibitori BET familije proteina JQ1 i I-BET762¹³⁵.

Među novim epigenetskim ciljnim mestima sa velikim potencijalom za razvoj novih grupa antitumorskih lekova ističu se sirtuin 2 (SIRT2) i WD-ponavljajući domen 5 protein (WDR5)^{136,137}. SIRT2 je predstavnik epigenetskih brisača, dok WDR5 ulazi u konstituisanje različitih multiproteinskih kompleksa sa funkcijom pisača i čitača epigenetskog koda¹³⁸. Inhibitori SIRT2 i WDR5 se još uvek nalaze u fazama pretkličkog razvoja i uprkos mnogim biološkim studijama kojima se potvrđuje antitumorski potencijal inhibicije ovih ciljnih mesta, još uvek nisu otkriveni kandidati koji ispunjavaju uslove za ulazak u kliničke studije^{139,140}. U daljem tekstu biće detaljnije obrazložen antitumorski potencijal, značaj rasvetljavanja mehanizama protein-ligand interakcije kao i značaj konformacione fleksibilnosti mesta interakcije za dalji razvoj inhibitora ova dva epigenetska proteina.

1.4.1. SIRT2

SIRT2 je NAD⁺-zavisna protein deacilaza zadužena za uklanjanje posttranslacionih modifikacija lizina¹⁴¹. SIRT2 kroz svoje interakcije sa velikim brojem citoplazmatskih i nuklearnih proteina ostvaruje ulogu regulatora velikog broja bioloških funkcija uključujući održavanje stabilnosti genoma, regulaciju metabolizma, starenja, tumorogeneze kao i ćelijskog ciklusa^{142,143}. Kao enzim koji se tradicionalno svrstava u histon-deacetilaze (klasa III histon deacetilaza), tzv. „brisače“ epigenetskog koda, SIRT2 se najčešće povezuje sa deacetilovanjem acetil-lizina histonskih proteina¹⁴⁴.

Ne treba zanemariti i značajnu ulogu SIRT2 u uklanjanju i ostalih acil-modifikacija lizina poput krotonil ostaka i dugolančanih acil modifikacija kao što su miristoil, 4-oksonanoil, dekanolil i heksanoil¹⁴⁵⁻¹⁵⁰. Acilovanje lizina predstavlja jednu od najzastupljenijih post-translacionih modifikacija kojima se, pored kondenzacije hromatina, regulišu i funkcije mnogih proteina unutar ćelija^{141,151-156}. Studije na ćelijskim i animalnim modelima utvrdili su veliki potencijal SIRT2 inhibicije u terapiji bolesti povezanih sa starošću, poput neurodegenerativnih bolesti ili karcinoma^{157,158}.

Ohrabrujući rezultati velikog broja pretkličkih istraživanja izvedenih u poslednjoj deceniji rezultovalo je u velikom interesovanju naučne javnosti ka razvoju SIRT2 inhibitora kao potencijalnih antineoplastika¹⁵⁸⁻¹⁶². Utvrđeno je da je inhibicija SIRT2 tesno povezana sa različitim aspektima razvoja i progresije malignih tumora, uključujući proliferaciju, invaziju, angiogenezu i metastatski potencijal^{158,162,163}. Dodatno, utvrđeno je da se SIRT2 uključuje i u mehanizme razvoja hemorezistencije malignog tumora na neke od poznatih hemioterapeutika. U tom kontekstu, najnovija istraživanja pokazala su sinergistički potencijal SIRT2 inhibitora u kombinaciji sa doksorubicinom, dasatinibom ili paklitakselom u tretmanu rezistentnog melanoma i specifičnih podtipova adenokarcinoma dojke¹⁶⁴⁻¹⁶⁶. O značaju SIRT2 kao obećavajućeg novog ciljnog mesta dejstva u savremenoj terapiji malignih tumora govore najnovija istraživanja usmerena ka primeni SIRT2 inhibitora u T-ćelijskoj imunoterapiji himernim antigenskim receptorom (eng. *Chimeric antigen receptor T-cell therapy* – CAR-T). Naime, SIRT2 inhibitori su pokazali veliki potencijal u aktivaciji tumor-infiltrirajućih limfocita što može rešiti i problem slabog odgovora na CAR-T¹⁶⁷. O značaju SIRT2 inhibicije na prevenciju najranijih faza karcinogeneze putem indukcije mitotičke katastrofe u ksenograft modelima tumora govore rezultati najnovijeg istraživanja kojim je ukazano i na hemopreventivni potencijal farmakološke inhibicije SIRT2¹⁶¹.

Ipak, antitumorski efekat povezan sa SIRT2 inhibitorima nije univerzalan i ne može se generalizovati na sve tipove malignih tumora i njihove stadijume^{158,168}. Jedan od najbolje izučavanih tipova malignih tumora osetljivih na SIRT2 inhibiciju jeste trostruko negativni adenokarcinom dojke (eng. *Triple-negative breast cancers* – TNBC)¹⁶⁹. SIRT2 inhibicija u TNBC

može dovesti do gubitka bazalnog diferencirajućeg fenotipa, gubitka invazivnosti i značajnog antiproliferativnog efekta što se može pripisati promociji degradacije c-MYC onkoproteina i Slug destabilizaciji^{158,162,170}. Naročito je interesantna povezanost inhibicije SIRT2 sa ubikvitinacijom i degradacijom c-MYC onkoproteina, za koji se već 40 godina bezuspešno radi na pronalaženju adekvatnog načina za farmakološku modulaciju¹⁶².

MYC onkoproteini predstavljaju familiju od tri proteina (c-, N- i L-MYC) koji kroz dimerizaciju sa MAX proteinom kreiraju DNA-vezujući transkripcioni faktor koji moduliše ekspresiju ciljnih gena povezanih sa rastom i proliferacijom¹⁷¹. Treba istaći da svoju funkciju transkripcionog faktora MYC-MAX postiže stupajući u direktnu, ili indirektnu interakciju sa velikim brojem proteina iz epigenetske mašinerije dovodeći do značajnih promena u organizaciji hromatina.^{172,173} Veliki broj ovih interakcija se dešava putem N-terminusa MYC onkoproteina zbog čega se MYC smatra i jednim od najbitnijih regulatora epigenoma¹⁷³.

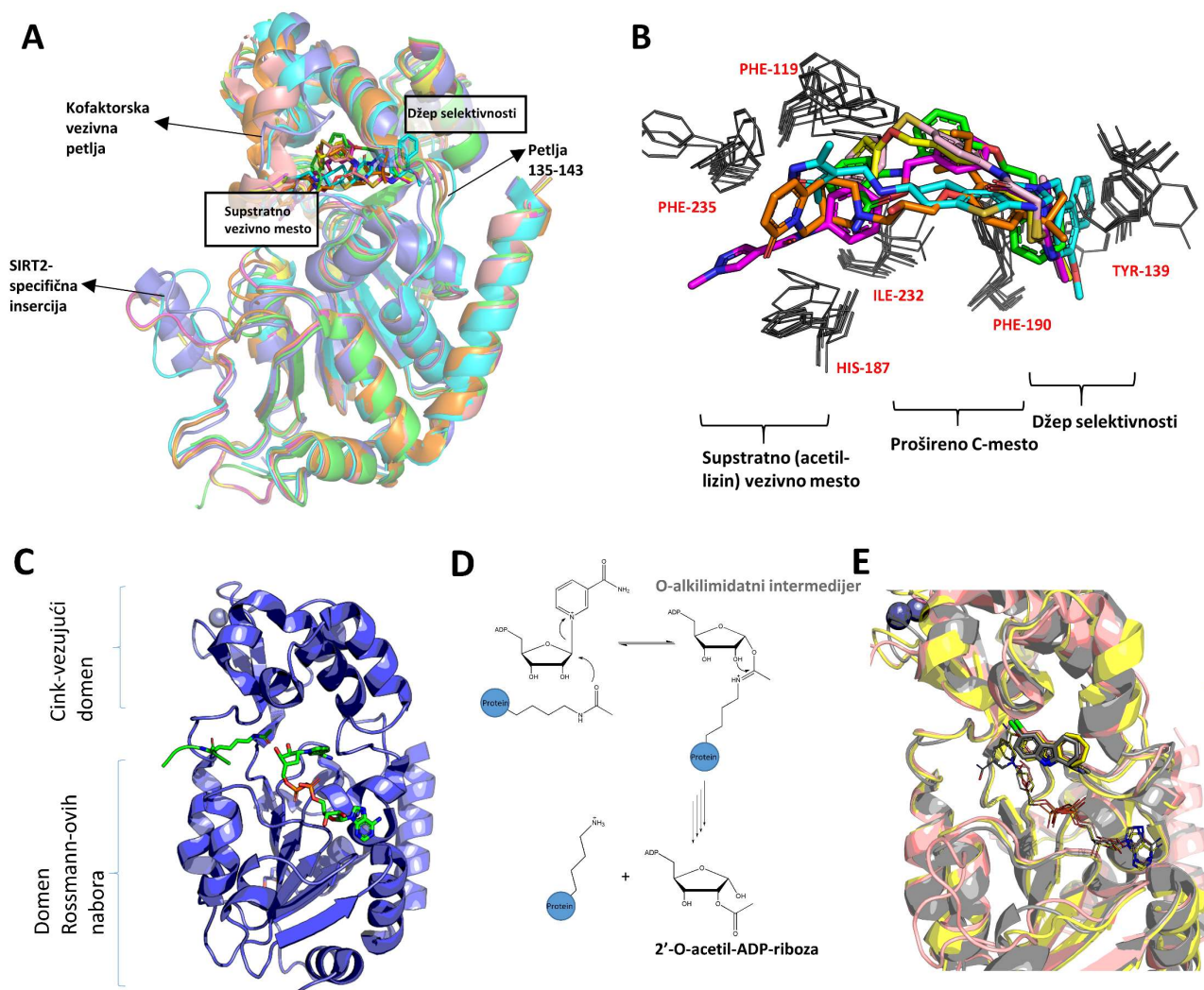
Prekomerna ekspresija MYC-a je opisana u gotovo polovini maligniteta i određena istraživanja ukazuju da se čak jedna trećina smrtnih slučajeva izazvanih malignitetom može povezati sa aberantnom ekspresijom, ili aktivnošću MYC onkoproteina. Brojni eksperimenti na animalnim modelima su ukazali da se inaktivacijom MYC-a postiže regresija tumora, čak i u slučajevima kada MYC ne pripada primarnoj onkogenoj leziji, zbog čega se MYC svrstava u grupu visokovalidiranih ciljnih mesta u terapiji malignih tumora. Sa druge strane, visoka intrizična strukturalna neuređenost MYC-a i nedostatak dobro struktuiranog vezivnog mesta na proteinu onemogućava razvoj antineoplastika za ovo ciljno mesto^{174,175}.

Nedavno okarakterisan SIRT2 inhibitor iz grupe peptidomimetika – TM pokazao je visoku selektivnost i potentnost ka SIRT2. TM ispoljava rasprostranjenu antitumorsku aktivnost koja koreliše sa SIRT2 inhibicijom i sa smanjenjem intracelularne koncentracije c-MYC onkoproteina, kada je u pitanju nekoliko ćelijskih linija adenokarcinoma dojke. Na mišjem modelu adenokarcinoma dojke, TM je pokazao značajnu inhibiciju rasta tumora uz relativno nisku toksičnost ka zdravim ćelijama¹⁶². Ovo istraživanje je jedno od poslednjih značajnih dokaza koji idu u prilog daljeg istraživanja inhibicije SIRT2 kao potencijalne strategije razvoja antitumorskih lekova i opravdava nastavak potrage za novim SIRT2 inhibitorima poboljšanih osobina.

Uprkos obećavajućim rezultatima, nijedan od SIRT2 inhibitora još uvek nije ušao u klinička istraživanja. Iako je do danas prijavljen veliki broj inhibitora, najveći broj njih je opterećen niskom potentnošću, neselektivnošću, ili lošim fizičko-hemijskim osobinama što ostavlja prostor za nastavak potrage za boljim SIRT2 inhibitorom^{151,163,176}. Treba istaći da je do danas na SIRT2 aktivnost profilisano preko 1000 jedinjenja što otvara prostor za uspostavljanje modela mašinskog učenja i uspostavljanje *in silico* platforme za dizajn zavisan od strukture liganda koja bi dodatno promovisala i ubrzala otkriće kandidata za kliničke studije.

1.4.1.a. Struktura aktivnog mesta SIRT2

Katalitičko jezgro SIRT2 se sastoji od većeg domena *Rossmann*-ovih nabora i manjeg cink-vezujućeg domena koji su međusobno povezani sa četiri fleksibilne petlje (Slika 1.14)¹⁷⁷. Na dodirnoj površini dva strukturalna domena oivičenog fleksibilnim petljama nalazi se aktivno mesto enzima¹⁷⁸. Aktivno mesto enzima predstavlja veliku i fleksibilnu šupljinu i može se podeliti na nekoliko segmenta: supstratno vezivno mesto (acetil-lizin vezujuće mesto), C-mesto, odnosno prošireno C-mesto i džep selektivnosti^{136,179}. U supstratno vezivno mesto se kao što i samo ime nalaže vezuje acilovana lizinska rezidua. U zavisnosti od dužine acil-modifikacije, acil-ostatak može da interaguje i sa ostalim delovima aktivnog mesta enzima. C-mesto je mesto vezivanja nikotinamida iz NAD⁺ koji je ključan za katalitički mehanizam, dok prošireno C-mesto predstavlja šupljinu iznad samog nikotinamida za koje se vezuju različiti supstrati, intermedijeri katalize (npr. slobodan nikotinamid), ili inhibitori^{179,180}. Džep selektivnosti predstavlja relativno novo otkriveno mesto koje nastaje konformacionom adaptacijom SIRT2 na specifičnu grupu selektivnih inhibitora od čega i potiče ime (Slika 1.14)¹⁸⁰.



Slika 1.14. Prikaz osnovne strukture, deacilacionog mehanizma i strukturne heterogenosti SIRT2. A – Prikaz poravnanih kristalografski razrešenih SIRT2:inhibitor kompleksa obrađenih u okviru disertacije (PDB ID:5Y0Z – narandžasta, PDB ID:5YQO – purpurna; PDB ID:5MAT – tirkizna; PDB ID:5YQL – žuta; PDB ID:5Y5N – zelena; PDB ID:5YQM – roze), sa naznačenim strukturnim elementima; B – Uvećanje vezivnih mesta inhibitora struktura na slici A sa naznačenim strukturnim elementima; C – Prikaz kompleksa SIRT2 sa supstratom i kofaktorom (zeleno) (PDB ID:4FVT) uz naznaku osnovnih domena; D – Prikaz katalitičkog mehanizma deacilovanja; E – Prikaz problematike dostizanja selektivnosti inhibitora kroz strukture SIRT1 (žuto) (PDB ID:4I5I), SIRT2 (roze) (PDB ID:5D7P) i SIRT3 (sivo) (PDB ID:4BV3) u kompleksu sa strukturno srodnim inhibitorima i NAD^+ /ADP-ribozom.

Ipak, džep selektivnosti nije usko vezan za konformacione promene indukovane selektivnim inhibitorima već se otvara i konformacionom adaptacijom SIRT2 na miristoilovani lizin¹⁸¹. Dakle, može se pretpostaviti da selektivnost inhibitora potiče od stabilizacije konformacionog stanja SIRT2 koje prirodno odgovara vezivanju dugolančanih supstrata.

Katalitički mehanizam deacilovanja započinje nukleofilnim napadom kiseonika karbonilne grupe acil-lizin supstrata na 1' ugljenik riboze iz NAD^+ , pri čemu nastaje O-alkilimidatni intermedijer i oslobađa se nikotinamid. U narednom koraku visoko konzervirana rezidua histidina deprotonuje 2'-OH riboze, koji zatim kroz intramolekulski nukleofilni napad na imidatni ugljenik formira biciklični intermedijer. Biciklični intermedijer potom hidrolizuje na lizin i 2'-O-acil-ADP-ribozu (Slika 1.14)
177,181,182

Većina SIRT2 inhibitora otkrivenih do danas interferira sa predstavljanim katalitičkim mehanizmom putem kompeticije sa supstratom i/ili kofaktorom (NAD⁺) u okviru aktivnog mesta enzima. Usled visoke konzerviranosti strukture vezivnog mesta između različitih izoformi familije sirtuina, naročito izoformi SIRT1 i SIRT3, postizanje selektivnosti inhibitora predstavlja jedan od najvećih izazova dizajna novih inhibitora. Neselektivni inhibitori često dele i gotovo identičan vezivni mod u okviru vezivnog mesta SIRT1-3 (Slika 1.14). Nedavno otkrivena farmakološka prednost selektivnih u odnosu na neselektivne SIRT2 inhibitore dodatno je pozicionirala selektivnost kao jedan od imperativa u razvoju novih inhibitora^{176,183,184}.

Strukturnim i mehanističkim studijama sprovedenim na SIRT2, utvrđeno je da vezivno mesto SIRT2 može da prihvati veći broj različitih acil ostataka lizina u rasponu od 2-14 C atoma (od acetil do miristoil ostatka). Promiskuitetnost SIRT2 prema različitim acil-ostacima može se pripisati konformacionoj fleksibilnosti SIRT2 enzima¹⁴⁶⁻¹⁵⁰. Kristalografske studije utvrdile su da je vezivno mesto SIRT2, odnosno mesto na samom proteinu za koje se vezuju ili supstrat, ili kompetitivni inhibitor, podložno konformacionim promenama. Naime, samo vezivno mesto izgrađeno je od petlji bez sekundarne strukturne uređenosti što uslovljava povećanu fleksibilnost ovog dela SIRT2²¹. Dodatno, kristalografske studije izvedene na SIRT2 u kompleksu sa strukturno različitim inhibitorima utvrdile su postojanje različitih konformacionih stanja SIRT2 u kompleksu sa hemijski raznorodnim grupama inhibitora (Slika 1.14). Ova strukturna heterogenost samog SIRT2 umnogome otežava racionalni dizajn novih inhibitora i ukazuje na značaj molekulske dinamičke razmatranja tokom dizajna novih predstavnika. Tokom godina rada na problematici racionalnog dizajna novih SIRT2 inhibitora, izvedene su i određene MD studije dinamike SIRT2:inhibitor kompleksa. Međutim, nijedno od prethodnih istraživanja nije za cilj imalo detaljnu karakterizaciju dinamike vezivnog mesta SIRT2, niti karakterizaciju svih dostupnih konformacionih stanja vezivnog mesta^{21,185-188}.

Pod pretpostavkom da dostupne X-ray strukture ne opisuju sva moguća stanja SIRT2 podobna za vezivanje inhibitora, opisivanje pune konformacione dinamike predstavlja informaciju od neprocenjivog značaja u dizajnu novih hemijskih grupa inhibitora primenom tehnika dizajna zavisnog od 3D-strukture ciljnog mesta.

1.4.2. WDR5

WDR5 je visoko konzervirani nuklearni protein koji se kroz izgradnju multiproteinskih kompleksa uključuje u regulaciju brojnih procesa unutar ćelije¹⁸⁹. Jedna od najznačajnijih uloga WDR5 predstavlja struktuiranje epigenetskih kompleksa „pisača“, preciznije MLL kompleksa histon metiltransferaza (HMTs), nespecifičnog letalnog (NSL) i Ada2-sadržavajućeg (ATAC) histon aciltransferaznog (HAT) kompleksa, uključenih u regulaciju remodelovanja hromatina. WDR5 takođe funkcioniše kao epigenetski „čitač“ koji prepoznaje specifične posttranslacione modifikacije na histonu H3^{138,190}. Poslednja istraživanja potvrdila su i njegovu ulogu kao ključnog kofaktora u interakciji MYC porodice onkoproteinskih transkripcionih faktora sa hromatinom^{175,191}.

WDR5 doprinosi karcinogenezi velikog broja malignih tumora. U akutnim leukemijama povezanim sa rearanžmanom MLL gena lociranog na hromozomu 11q23, WDR5 se uključuje u leukemogenezu kroz kompleksan mehanizam interakcija između divljeg tipa (eng. *wilde-type*) MLL kompleksa i fuzionih MLL onkoproteina. Kroz fuziju sa različitim genima od kojih su neki transkripcioni aktivatori, nastaju fuzionih MLL proteini bez C-terminalnog SET domena, a samim tim i bez metiltransferazne aktivnosti na H3K4. Fuzionih MLL protein stupa u interakciju sa divljim tipom MLL kompleksa i, iskorišćavajući sposobnost interakcije MLL kompleksa sa H3K4, dovodi do aktivacija transkripcije Hox i Meis1 gena i posledično leukemogeneze^{174,192-194}.

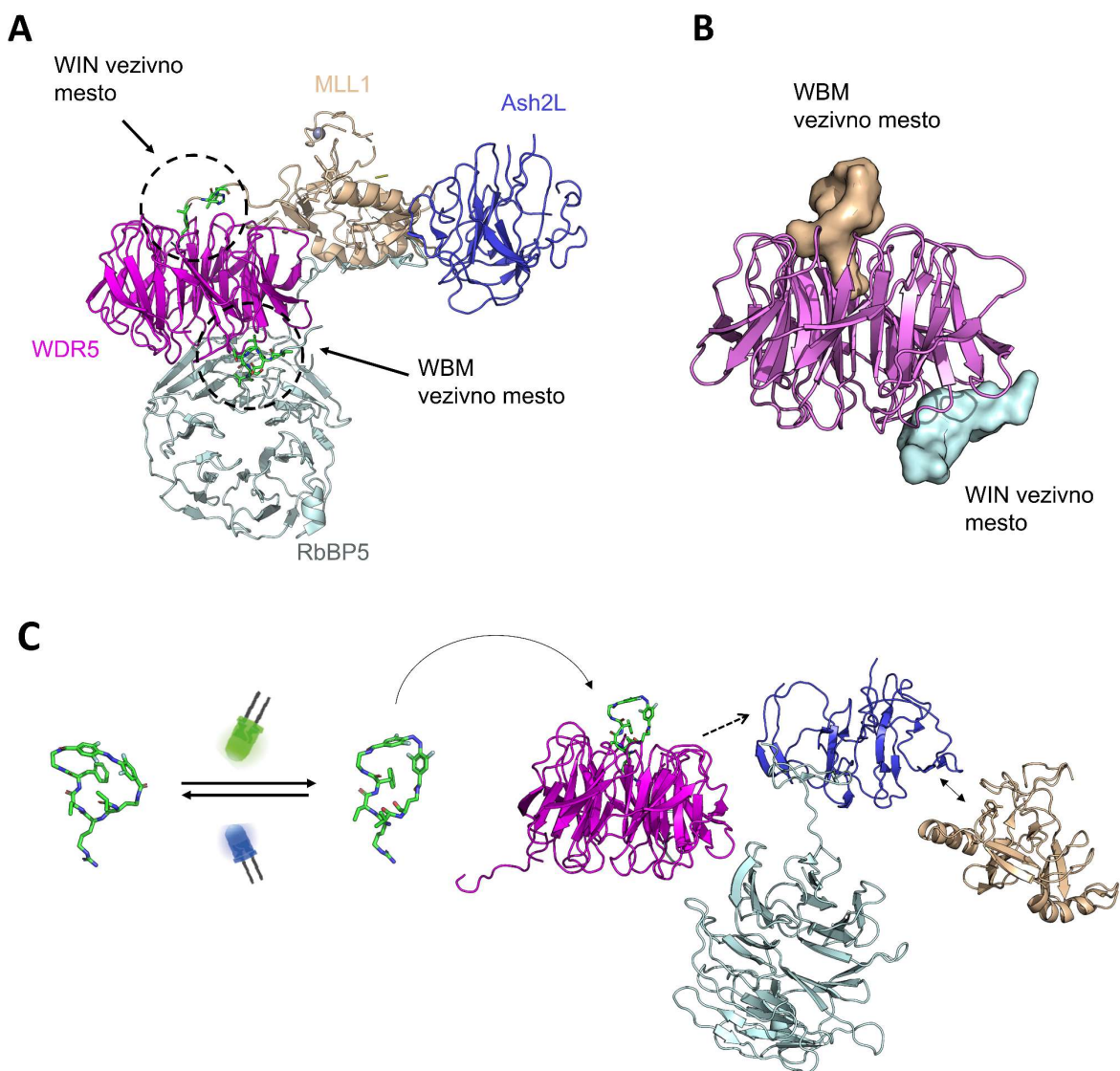
Sa druge strane, prekomerna ekspresija WDR5 je povezana i sa lošom prognozom velikog broja agresivnih solidnih tumora, poput kolorektalnog karcinoma, karcinoma pankreasa, dojke, jetre, neuroblastoma itd.¹⁹⁵ Pored klasične uloge WDR5 u struktuiranju epigenetskih kompleksa bitnih za epigenetske mehanizme karcinogeneze i progresije malignih tumora, nedavno je ovaj protein

okarakterisan i kao ključni kofaktor u MYC-promovisanoj tumorogenezi solidnih tumora^{174,175,196}. WDR5 kroz direktnu interakciju sa MYC omogućava asocijaciju MYC-a za hromatin i na taj način dovodi do transkripcione aktivacije ciljnih gena bitnih za tumorogenezu. Razvoj malih molekula inhibitora protein-protein interakcije WDR5 danas predstavlja jednu od strategija za indirektno ciljanje MYC onkoproteina i razvoj novih klasa antitumorskih lekova¹⁹⁵.

1.4.2.a. Struktura WDR5

Studije strukturne biologije utvrdile su da se WDR5 sastoji od sedam tipičnih WD40 ponavljajućih domena koji formiraju sedam lopatica u okviru beta-propeler strukture. Svaka lopatica je sastavljena od četiri standardna antiparalelna beta-lanca (Slika 1.15). Aksijalna šupljina u strukturi WDR5 predstavlja peptidno vezivno mesto u okviru koga se ostvaruju protein-protein interakcija sa interakcionim partnerima poput histona H3, ili SET domena MLL H3K4 metiltransferaza. U okviru ovog vezivnog mesta dolazi do visokoafinitetnog prepoznavanja između WDR5 i arginina-2 histona 3, odnosno WIN motiva SET domena metiltransferaza¹⁹⁷. Zajedničko za obe interakcione peptidne sekvence jeste prisustvo arginina koji ulazi direktno u centralnu šupljinu, dok ostatak peptida ostvaruje interakciju sa delom vezivnog mesta smeštenim na površini WDR5 (Slika 1.15). Studije su ukazale na strukturnu plastičnost površinskog dela vezivnog mesta WDR5 što za posledicu ima sposobnost WDR5 da se prilagodi većem broju različitih peptida koji u osnovi posedju WIN motiv.¹⁹⁸

Na suprotnoj strani arginin-vezujućeg vezivnog mesta WDR5, smešteno je tzv. WBM mesto za koje se vezuju proteini koji u svojoj strukturi posedjuju WBM motiv (eng. WDR5-binding motif), poput MYC onkoproteina, ili RBBP5.¹⁹⁰ Strukturne studije su ukazale da WDR5 pri formiranju stabilnih multiproteinskih kompleksa ostvaruje istovremenu protein-protein interakciju putem oba mesta (WIN i WBM) sa interakcionim partnerima¹⁹⁹. Iako su strukture WDR5 sa kratkim peptidnim sekvencama navedenih motiva razrešene na visokoj rezoluciji primenom kristalografije X-zraka i strukture multiproteinskih kompleksa razrešene na nešto nižoj rezoluciji primenom krio-elektronske mikroskopije, dinamički aspekti ove interakcije do danas ostaju nedovoljno istraženi (Slika 1.15).



Slika 1.15. Osnovni prikaz WDR5 u okviru MLL1 kompleksa. A – Struktura MLL1 kompleksa (PDB ID: 6KIU) sa osnovnim protein-protein interakcijama u okviru WDR5 označenim zelenom bojom; B – Uvećanje WBM i WIN vezivnog mesta sa naznačenim površinama; C – Princip rada optofarmakoloških inhibitora WDR5 protein-protein interakcije na primeru MLL1 kompleksa.

Nasuprot SIRT2, koji poseduje aktivno mesto u vidu voluminozne šupljine, aktivna mesta na WDR5 predstavlja zapravo dodirne površine kojom WDR5 ostvaruje svoju uloga u konstituisanju multiproteinskih kompleksa (Slika 1.15).

Razvoj inhibitora protein-protein interakcija (PPI) WDR5 je primarno fokusiran na arginin-vezujuće (WIN) mesto, koje ima jasnije definisanu strukturu sa jednom centralnom šupljinom za vezivanje arginina što predstavlja odličnu polaznu tačku za dizajn novih liganada. Zaravnjena struktura WBM mesta i nedostatak jasno definisanih šupljina čine dizajn liganada za WBM mesto veoma komplikovanim. Iako su do danas otkriveni neki ligandi za WBM mesto, razvoj WIN inhibitora ostaje glavna struja u razvoju inhibitora WDR5^{190,195}. Zbog činjenice da je za postizanje funkcionalnih multiproteinskih kompleksa neophodna interakcija putem oba mesta na WDR5, WIN inhibitori postižu slične efekte poput WBM inhibitora¹⁹⁵. Usled nedovoljnog poznavanja dinamičkih osobina WDR5, do danas ostaje nedovoljno razjašnjeno da li navedena mesta međusobno alosterno komuniciraju²⁰⁰. Otkriće alosterne komunikacija između ova dva mesta može pružiti značajne smernice u dizajnu efikasnijih inhibitor PPI WDR5.

Istorijski gledano, razvoj inhibitora PPI je dugo posmatran kao nemoguć pre svega zbog velike površine na kojoj se odigrava PPI i njene konformacione fleksibilnosti. Međutim, napredak metoda racionalnog dizajna, kao i uspeh nekoliko kliničkih kandidata, opovrgnuli su ovaj tradicionalni stav^{201–203}. Usled velike površine na kojoj se odigrava PPI, inhibitori protein-protein interakcija veoma često predstavljaju peptide, odnosno peptidomimetike/ciklične peptide molekulske mase iznad 500 Da²⁰⁴. Zbog fleksibilnosti vezivnih površina PPI i postojanja velikog broja prolaznih vezivnih tačaka na toj površini, kao i zbog fleksibilnosti samih peptidnih inhibitora i postojanja potencijalno većeg broja vezivnih modova inhibitora, racionalni dizajn peptidomimetskih inhibitora ostaje izazovna oblast u kojoj simulacije molekulske dinamike predstavljaju jedan od neizostavnih alata^{71,201,203–205}.

Iako su WDR5 inhibitori u ranoj fazi razvoja i još uvek nema kliničkih kandidata, interesovanje ka WDR5 inhibitorima kao novim antineoplasticisma ne jenjava što potvrđuje i veliki broj istraživanja objavljenih poslednjih par godina^{137,138,206–211}. Posebnu pažnju izaziva pojava optofarmakoloških peptidomimetskih inhibitora WDR5 (Slika 1.15)^{212–214}. Optofarmakološki inhibitori predstavljaju grupu fotoreaktivnih inhibitora WDR5 koji prilikom ozračivanja svetlošću određene talasne dužine zauzimaju farmakološki aktivna, ili inaktivna konformaciona stanja. Ovakvi molekuli omogućavaju postizanje visokog stepena kontrole u vremenu trajanja efekta leka kao i u lokalizaciji efekta u odgovarajuće delove organizma koji su izloženi elektromagnetnom zračenju određene talasne dužine. Za postizanje svetlošću-modulisanog odgovora neophodno je uvođenje fotoreaktivnih grupa u strukture molekula. Azobenzeni predstavljaju jedne od najčešće upotrebljivanih fotoreaktivnih grupa²¹⁵.

Usled napretka metoda sinteze peptida na čvrstoj fazi, inkorporiranje azobenzenkih fragmenata u strukture peptida/peptidomimetika postala je atraktivna strategija dizajna optofarmakoloških inhibitora PPI. U nedavno objavljenoj studiji, po prvi put je okarakterisana svetlošću kontrolisana inhibicija PPI WDR5 *cis/trans* izomerizacijom azobenzenkog fragmenta u okviru peptidnog/peptidomimetskog inhibitora²¹². Uprkos skromnoj selektivnosti između *cis* i *trans* stanja inhibitora, ova studija je otvorila vrata ka razvoju nove grupe optofarmakoloških inhibitora WDR5. Primarni cilj daljeg razvoja ove grupe WDR5 inhibitora ostaje povećanje odnosa selektivnosti *cis* i *trans* izomera²¹³.

1.5. Primena bioinformatičkih analiza u otkriću lekova

Bioinformatika podrazumeva primenu računarskih metoda u upravljanju, analizi i interpretaciji bioloških podataka, uključujući genomičke, proteomske i metabolomske podatke. Sa pojavom tehnologija visokog protoka, kao što su sekvenciranje naredne generacije i masena spektrometrija, danas je moguće generisati ogromne količine podataka iz bioloških uzoraka²¹⁶. Bioinformatički alati i algoritmi mogu se koristiti za analizu i interpretaciju ovih podataka čime se doprinosi i boljem razumevanju molekularnih mehanizama bolesti, ili određene terapije. Bioinformatička analiza se u sve većoj meri koristi u personalizaciji terapije pacijenata i smatra se jednom od vodećih tehnika u nadolazećoj eri personalizovane medicine²¹⁷. Analizom genomskih i proteomskih podataka populacija pacijenata, moguće je identifikovati biomarkere za predviđanje odgovora na terapiju pojedinačnog pacijenta²¹⁸. Izučavanjem kompleksnih mreža intermolekulskih interakcija unutar ćelije, kao i transfera informacija unutar komunikacionih puteva, moguće je identifikovati neke od ključnih proteina uključenih u razvoj i progresiju bolesti, kao i odgovora na terapiju²¹⁹.

Otkrivanje i validacija novih ciljnih mesta je ključan korak u procesu razvoju lekova, kojim se postavlja temelj za razvijanje novih farmakoloških grupa. U kontekstu savremenog otkrića lekova, bioinformatička analiza velikih baza podataka igra jednu od ključnih uloga u identifikaciji novih ciljnih mesta za razvoj potencijalno efikasnijih terapeutika²¹⁹. S obzirom na veliku kompleksnost i nedovoljnu razjašnjenost mehanizama koji stoje u pozadini malignih i neurodegenerativnih bolesti, razvoj efikasnijih terapeutika se u sve većoj meri oslanja na polifarmakološki pristup. Razvoj jednog leka koji bi interagovao sa većim brojem ciljnih mesta od značaja za određenu bolest može

se smatrati poželjnim u slučaju jasno validiranih ciljnih mesta. Ipak, treba istaći da identifikacija novih ciljnih mesta predstavlja kompleksan zadatak²²⁰.

U slučaju malignih oboljenja, javno je dostupan određen broj baza u kojima su pohranjene velike količine podataka generisanih visoko propusnim metodama²²¹. Baze podataka o osetljivosti tumorskih ćelija na odgovarajuće lekove, ili hemijske probe, predstavljaju jedan od značajnih resursa za identifikaciju novih biomarkera odgovora na terapiju, ali i generisanja računarskih modela za predviđanje odgovora na kombinacije lekova²²²⁻²²⁵. S obzirom na činjenicu da veliki broj bioaktivnih jedinjenja poseduje poznate, ili nepoznate aktivnosti na sporednim ciljnim mestima (tzv. *off-target*), integracija podataka o osetljivosti tumorskih ćelija na terapiju sa transkriptomskim podacima može pružiti veoma korisne smernice u dizajnu polifarmakoloških liganada kao i u otkriću novih ciljnih mesta.

2. Ciljevi istraživanja

Jedan od glavnih ciljeva ove teze bio je ispitivanje konformacione dinamike proteina iz grupe epigenetskih ciljnih mesta dejstva (SIRT2 i WDR5) kod kojih konformaciona fleksibilnost samog proteina i/ili liganda predstavlja poduhvat sa aspekta racionalnog dizajna novih liganada. Razvojem novih protokola za računarsko dizajniranje lekova kroz implementaciju informacija iz velikih baza podataka i/ili molekulskih simulacija, kao i njihovom ekperimentalnom validacijom primenom organskih sinteza i *in vitro* bioloških ispitivanja, istraživanja sprovedena u okviru ove teze imala su za cilj da postave osnovu za budući razvoj novih hemijskih grupa inhibitora navedenih epigenetskih proteina kao potencijalno novih antineoplastika. Jedan od dodatnih ciljeva bio je i postavljanje smernica za budući razvoj dualnih inhibitora histon-deacetilaza (HDAC) u terapiji adenokarcinoma pankreasa kroz razvoj bioinformatičkog protokola za predviđanje sinergizama između bioaktivnih jedinjenja.

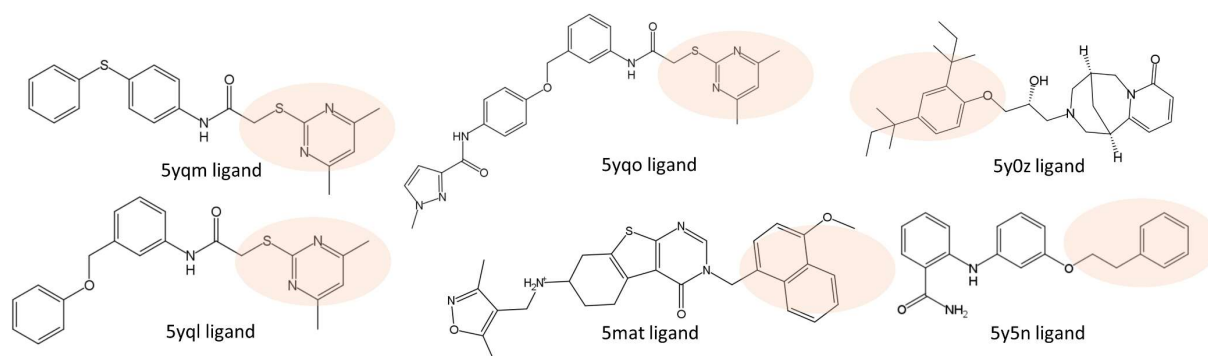
Specifični ciljevi istraživanja:

- Razvoj protokola biomolekulskih simulacija za detaljno opisivanje konformacione dinamike SIRT2 sa posebnim osvrtom na otkriće alternativnih konformacionih stanja SIRT2 i skrivenih džepova unutar vezivnog mesta SIRT2 inhibitora;
- Razvoj i eksperimentalna validacija SBVS modela za otkriće selektivnih SIRT2 inhibitora zasnovanih na alternativnim konformacionim stanjima SIRT2 opisanih kroz razvijeni protokol simulacija uz poseban osvrt na ispitivanje mogućnosti dosezanja novih oblasti hemijskog prostora SIRT2 inhibitora;
- Razvoj i validacija *in silico* platforme za racionalni dizajn novih selektivnih SIRT2 inhibitora zasnovane na visoko validiranim modelima kvantitativnog odnosa strukture i aktivnosti/selektivnost generisanih savremenim algoritmima mašinskog učenja pomoću velikih setova podataka;
- Sinteza i *in vitro* ispitivanja novih hemijskih vrsta SIRT2 inhibitora dizajniranih na osnovu rezultata prospektivnog virtuelnog skrininga, kao i primenom razvijenih modela odnosa strukture i aktivnosti/selektivnosti;
- Karakterizacija konformacionog ponašanja WDR5 u prisustvu različitih optofarmakoloških peptidomimetskih liganada primenom metoda molekulske dinamike, kao i postavljanje protokola za dizajn novih optofarmakoloških liganada zasnovanog na proračunima slobodne energije vezivanja;
- Kvantifikovanje energetske promene vezanih za proces fotoizomerizacije strukture azobenzena u okviru ispitivanog peptidomimetikog inhibitora WDR5, kao ključnog parametara opisa selektivnosti *cis* i *trans* izomera primenom metoda metadinamičkih simulacija;
- Uspostavljanje modela alosterne komunikacije dva vezivana mesta u okviru WDR5 na osnovu analize dinamičke mreže interakcija aminokiselinskih rezidua WDR5 dobijenih simulacijama molekulske dinamike;
- Razvoj bioinformatičkog protokola za predviđanje sinergizama između inhibitora familije histon-deacetilaza (uključujući sirtuine) i poznatih bioaktivnih jedinjenja u tretmanu adenokarcinoma pankreasa, kao i identifikacija novih ciljnih mesta za razvoj inovativnih dualnih inhibitora histon-deacetilaza.

3. Materijali i metode

3.1. Racionalni dizajn SIRT2 inhibitora

Ispitivanje konformacione dinamike sprovedeno je na uzorku od 7 inicijalnih PDB struktura (PDB ID:5MAT48, PDB ID:5Y5N49, PDB ID:5YQO50, PDB ID:5YQL50, PDB ID:5YQM50, PDB ID:5Y0Z51 i PDB ID:3ZGO58).



Slika 3.1. Pregled osnovnih SIRT2 inhibitora uključenih u simulacije. Zaokruženi delovi struktura interaguju sa džepom selektivnosti SIRT2.

3.1.1. Generisanje parametara polja sile

Za svaki od proučavanih inhibitora SIRT2 (ligand 5mat iz PDB ID:5MAT²²⁶; ligand 5y5n iz PDB ID:5Y5N²²⁷; ligand 5yqo iz PDB ID:5YQO²²⁸; ligand 5yql iz PDB ID:5YQL50²²⁸; ligand 5yqm iz PDB ID:5YQM²²⁸; ligand 5y0z iz PDB ID:5Y0Z¹⁴⁵) (Slika 3.1), topologije i parametri su najpre generisani upotrebom programa CGenFF (CHARMM General Force Field), verzija 2.4.0, kome je pristupljeno putem web-servera (cgenff.umaryland.edu)^{229,230}. U skladu sa konceptom prenosivosti CHARMM porodice sila polja, CGenFF dodeljuje parametre analognim postojećem skupu parametara. Loša analogija sa postojećim parametrima je naznačena vrednostima kaznenih skorova koje generiše sam server. Skorovi između 10 i 50 zahtevaju osnovnu validaciju parametara, dok vrednosti preko 50 zahtevaju ponovnu parametrizaciju. Sva naelektrisanja i nevezivni parametri sa kaznenim skorom većim od 10 su uzeti u obzir za optimizaciju korišćenjem *Force Field Toolkit* (ffTK) protokola³¹. Sve pripreme ulaznih QM datoteka i podešavanje naelektrisanja i vezivnih parametara su obavljene korišćenjem ffTK VMD dodatka^{31,231}. QM izračunavanja su izvedena korišćenjem ORCA 4.2.1^{232,233}, dok su molekularno-mehanička (MM) izračunavanja obavljena korišćenjem NAMD 2.13²³. U skladu sa standardnom CHARMM procedurom, većina potrebnih QM izračunavanja (minimalizacija, *Hessian*-ova matrica, skeniranje torzija) su obavljene na MP2 (6-31G*) nivou teorije. Budući da su QM izračunavanja na ovom nivou teorije za molekule veće od 40 atoma praktično neizvodljiva, pristup fragmentacije je korišćen u skladu sa opštom preporukom u literaturi. Ukratko, tačka fragmentacije je odabrana tako da se sačuvaju hemijske osobine grupa koja se optimizuje i da se obezbede fragmenti sa celobrojnim naelektrisanjem. Ovaj pristup fragmentacije je takođe korišćen za izbegavanje mogućih sternih sudara prilikom postavljanja molekula vode u optimizaciji parcijalnih naelektrisanja^{25,30,234–236}.

Fitovanje parcijalnih naelektrisanja izvršeno je prema opštim preporukama za CHARMM polje sile, koristeći interakciju sa TIP3P molekulima vode. U prvom koraku su SIRT2 inhibitori, ili njihovi fragmenti minimizovani, a zatim su molekuli vode postavljeni oko svakog liganda u orijentacijama koje maksimiziraju interakcije donora i/ili akceptora vodonične veze. U sledećem koraku energije interakcije liganada sa molekulima vode su minimizovane na HF(6-31G*) nivou teorije. Za ne-sp³ ugljenikove atome, računane su interakcione energije vode u oba pravca – razmatrajući interakcije donora i akceptora vodonične veze. Interakcija sa nižom energijom je korišćena za prilagođavanje naelektrisanja. Alifatičnim vodonocima je po konvenciji dodeljeno naelektrisanje +0.09e (kako je predviđeno na web-serveru). S obzirom da su QM proračuni izvedeni u gasnoj fazi, energetske

interakcije nenaektisanih molekula su skalirane faktorom 1.16 kako bi se kompenzovali efekti polarizacije u vodenom rastvoru. Skaliranje energije za naelektrisane molekule nije primenjivano. Parcijalna naelektrisanja sa najboljim slaganjem QM proračunatih i MM optimizovanih interakcionih energija i distanci sa TIP3P molekulima vode, korišćena su u narednoj optimizaciji vezivnih parametara polja sile. Konačna naelektrisanja celih molekula su dobijena spajanjem fragmenata prema ranije opisanim strategijama (naelektrisanja na rubnim atomima su izmenjena pripajanjem naelektrisanja uklonjenih vodonika)^{25,30,237}.

Veze i uglovi su optimizovani računanjem QM Hessian matrice internih koordinata na MP2(6-31G*) nivou teorije. Nakon što su veze i uglovi optimizovani, izvedeno je skeniranje relaksacije diedarskih uglova na MP2(6-31G*) nivou teorije. Diedarski uglovi unutar heterociklusa rotirani su u koracima od $\pm 30^\circ$ u rasponu od 5° , dok su ostali optimizovani diedarski uglovi rotirani u koracima od $\pm 180^\circ$ u rasponu od 15° .

3.1.2. Klasične MD simulacije

Klasična MD simulacija je izvedena za sledeće sisteme: SIRT2:5y5n (iz PDB ID:5Y5N²²⁷); SIRT2:5mat (iz PDB ID:5MAT²²⁶); SIRT2:5yqo (iz PDB ID:5YQO²²⁸); SIRT2:5yql (iz PDB ID:5YQL²²⁸); SIRT2:5yqm (iz PDB ID:5YQM²²⁸); SIRT2:5y0z (iz PDB ID:5Y0Z¹⁴⁵); SIRT2:apo (iz PDB ID:3ZGO¹⁸²). Nedostajuće rezidue su dodate korišćenjem MODELLER 9.21²³⁸. Početne strukture su protonovane, dok su tautomerni oblici ostataka određeni korišćenjem PROPKA programa, implementiranog u okviru serverske verzije PDB2PQR 3.1.0 softvera (<https://server.poissonboltzmann.org/pdb2pqr>)²³⁹⁻²⁴¹. Dominantni oblik na fiziološkoj pH pojedinačnih liganda je utvrđen u skladu sa pKa proračunima izvedenim u Marvin 21.3 programu²⁴², dok je dominantni oblik liganda unutar vezivnog mesta predviđen korišćenjem PROPKA 3.1 softvera. Usled rezultata predviđanja pKa u okviru PROPKA 3.1 softvera koji su ukazali na značajnu zastupljenost protonovanog stanja liganda 5y0z unutar vezivnog mesta SIRT2 (predviđeni pKa = 7,55), ligand 5y0z je u kompleksu sa SIRT2 simuliran u oba oblika (protonovanom – sistem SIRT2:5y0z_charged; molekulskom – sistem SIRT2:5y0z). Sve molekule vode prisutne u PDB strukturama su zadržane. MD simulacije su izvršene upotrebom GROMACS paketa (verzija 2020.4)²⁴³ i CHARMM36 polja sile²⁴⁴ sa TIP3P eksplicitnim modelom vode. Topologije i parametri liganda su pripremljene kao što je opisano u prethodnom odeljku. Cistein koordiniran sa Zn²⁺ je parametrizovan kao CYN ostatak. Svaki sistem je simuliran u oktaedarskoj periodičnoj kutiji. Naelektrisanje protein-ligand kompleksa je neutralizovano korišćenjem 0,15M NaCl. Particle-mesh Ewald (PME) algoritam je korišćen za elektrostatičke interakcije na dugom rastojanju, a Coulomb-ove i van der Waals-ove interakcije su računane za rastojanja od maksimum 12 Å. Sve veze sa H-atomima su konvertovane u ograničenja korišćenjem LINCS algoritma²⁴⁵. Za integraciju jednačine kretanja upotrebljen je leapfrog algoritam²⁰. Sistemi su početno minimizirani korišćenjem metode gradijenta strmog spusta.

Ekvibracija je izvedena u NVT ansamblu (temperatura od 310 K, V-rescale termostat), a zatim u NPT ansamblu (pritisak od 1 atm, Parinello-Rahman barostat) uz korišćenje pozicionih ograničenja na proteinima i ligandima. Poziciona ograničenja su postepeno uklanjana korišćenjem konstanti sile od 1000, 100 i 10 kJ/mol nm⁻² tokom 1 ns NPT ekvibracije. Produkcijske simulacije su izvedene u NPT ansamblu (temperatura od 310 K i Nose-Hoover termostat, pritisak od 1 atm i Parinello-Rahman barostat). Ukupno 200 ns produkcijskih simulacija u tri replike je izvedeno za svaki simulirani sistem. Analiza trajektorija (koren srednjeg kvadratnog odstupanja (eng. *Root-mean-square deviation* – RMSD), koren srednjeg kvadratne fluktuacije atomskih pozicija (eng. *Root-mean square fluctuation* – RMSF) rezidua, analiza glavnih komponenti (eng. *Principal component analysis* – PCA)) izvedena je korišćenjem MDAnalysis 2.0.0²⁴⁶ i GROMACS softvera (gmx covar i gmx anaeig). Mreže interakcija između rezidua su dobijene iz trajektorija korišćenjem softvera gRINN 1.1.0²⁴⁷. Računanje metrike prespajanja dobijenih mreža izvedeno je primenom DyNet 1.0.0 dodatka u okviru softvera Cytoscape 3.7.2^{248,249}. Trajektorije su klasterovane primenom gromos metode klasterovanja implementirane unutar GROMACS 2020.4 programa²⁵⁰.

3.1.3. Metadinamička simulacija

Analiza nezavisnih komponenti sa vremenskim kašnjenjem (eng. *Time-Lagged Independent Component Analysis* – tICA) je metoda za obradu signala kojom se redukuje dimenzionalnost podataka uz očuvanje kinetičkih podataka^{251,252}. Metoda je našla veliku primenu u analizi dinamike proteina gde se koristi za identifikaciju i analizu sporih pokreta proteina. tICA identifikuje spore nezavisne kolektivne stepene slobode analizirajući matriks kovarijanse vremenskog kašnjenja $C(\tau)$ (jednačina 25) u kom x može predstavljati atomsku poziciju u trajektoriji, ili drugu veličinu izvedenu iz trajektorije, dok je τ vreme kašnjenja. Rešenje tICA problema (jednačina 25) predstavlja projekcioni vektor koji maksimizuje autokorelacionu funkciju koji se može dobiti rešavanjem klasičnog generalizovanog problema svojstvenih vrednosti (eng. *eigenvalue*) prikazanog jednačinom (26), gde v_k predstavlja traženi svojstveni vektor (eng. *eigenvector*), dok λ_k predstavlja sopstvenu vrednost vektora²⁵³. tICA analiza je upotrebljena za definisanje kolektivnih varijabli u metadinamičkoj simulaciji.

$$C(\tau) = ((x_i(t)x_j(t + \tau)_t))_{ij \in \mathbb{R}^{d \times d}} \quad (25)$$

$$C(\tau)v_k = \lambda_k C(0)v_k \quad (26)$$

Dobro temperirana metadinamička simulacija (MetaD) je izvršena prema opštim preporukama koje su predložili Pande i saradnici²⁵⁴. Prvo, pomoću MSMBuild 3.9.0 alata²⁵⁵, izračunati su diedarski uglovi ϕ i ψ aminokiselinskih ostataka unutar vezivnog mesta za skup trajektora dobijenih kroz klasične MD simulacije. Koristeći isti alat, izvršena je tICA analiza^{251,252} sa vremenom kašnjenja od 10 ns. Simulacije dobro temperirane metadinamike (MetaD) izvedene su pomoću PLUMED 2.6.4 programa povezanim sa GROMACS 2020.4 programom^{256,257}. Za MetaD simulacije odabrana su tri sistema: SIRT2:apo, SIRT2:5y5n i SIRT2:5yqo. Sistemi su pripremljeni na isti način kao i za klasičnu MD. Prve dve nezavisne komponente (tIC-a) su konvertovane u PLUMED format prema kodu koji su obezbedili Pande i saradnici²⁵⁴. Metadinamičke simulacije izvedene su sa širinama *Gaussian*-ovih potencijala podešenim na 0,2. *Gaussian*-ovi potencijali pristrasnosti su depozitovani svakih 1000 koraka sa početnom visinom od 1 kcal/mol i faktorom pristrasnosti 50. Svi sistemi su simulirani dok visina depozitovanih Gausovih potencijala nije pala ispod 0,1 (10% početne visine) i profili slobodne energije ostali bez većih fluktuacija najmanje 100 ns simulacije (približno 700-800 ns simulacija za svaki sistem)²⁵⁸⁻²⁶¹. Nakon uklanjanja zidnog iskrivljenja, dobijene MetaD trajektorije su ponderisane i projektovane na dve CV (CV1 - diedar chi1 ASP-170; CV2 - rastojanje između (CZ)PHE-190 i (CB)ALA-135). Selekcija relevantnih konformacionih stanja za virtuelni skrining izvršena je korišćenjem dvostepenog postupka. U prvom koraku, oko 20% stanja sa minimalnom energijom na tIC1 vs tIC2 površini je izdvojeno. U drugom koraku, ovi podskupovi konformacionih stanja su analizirani na CV1/CV2 površini i oni koji su se nalazili u energetske minimumima na CV1/CV2 površini odabrani su za retrospektivnu VS analizu.

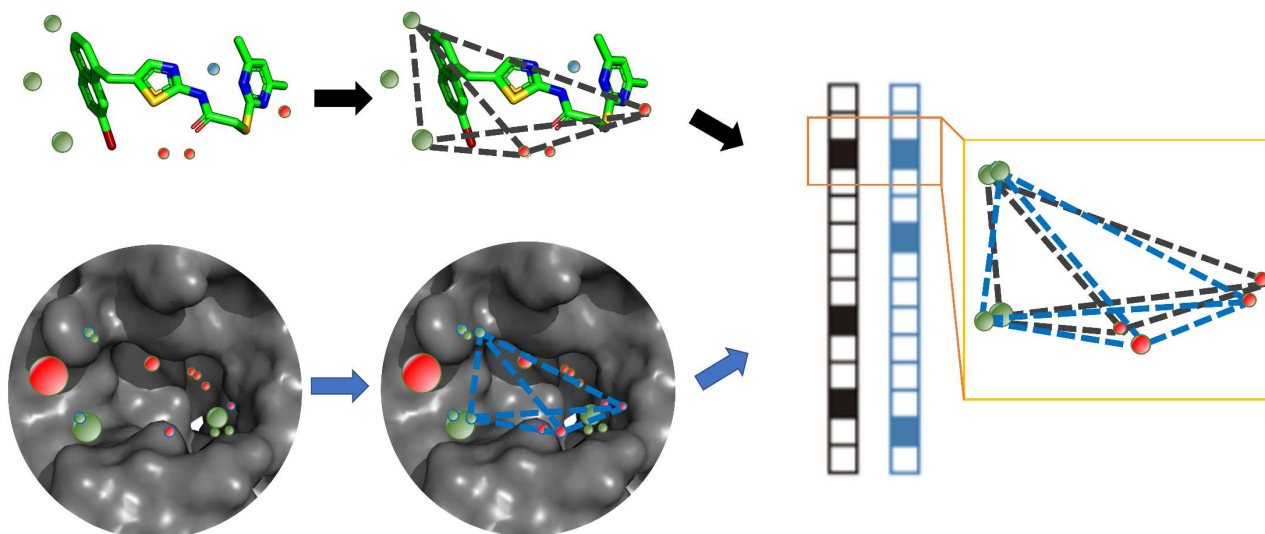
3.1.4. SBVS

Za SBVS korišćen je softver FLAP 2.2.1 (eng. *Fingerprints for Ligands and Proteins*)^{68,262,263}. FLAP se oslanja na metodologiju opisivanja malih molekula i proteina kroz četvorokomponentni farmakoforni fingerprint, izračunatih iz molekulskih polja interakcije (eng. *Molecular Interaction Fields* – MIF) pomoću GRID programa. Inicijalno GRID opisuje različite MIF-ove kroz računanje energije interakcije (E_{xyz}) između odgovarajuće hemijske probe i molekula u svakoj tački 3D prostora koristeći molekulsko-mehaničku formulaciju predstavljenu jednačinom (27):

$$E_{xyz} = \sum_{i=1}^n E_{LJ} + \sum_{i=1}^n E_{el} + \sum_{i=1}^n E_{hb} + S \quad (27)$$

gde su E_{LJ} Lennard-Jones-ov potencijal, E_{el} elektrostatski potencijal, E_{hb} potencijal vodoničnog vezivanja i S entropijski faktor²⁶⁴. Najčešće korišćene hemijske probe su: proba H - opisuje oblik ispitivanog molekula; proba O (sp^2 karbonilni kiseonika vezan za 1 susedni atom i sa 2 slobodna

elektronska para) – opisuje donorske H-veze molekula; proba N1 (planarni azot povezan sa jednim vodonikom i dva susedna atoma, bez slobodnih elektronskih parova) – opisuje akceptorske H-veze ispitivanog molekula; DRY proba – procenjuje hidrofobne osobine ispitivanog molekula. Kompleksnost MIF-ova opisanih oko molekula se dalje redukuje odabirom reprezentativnih tačaka primenom odgovarajućih funkcija koje vrednuju volumen odgovarajućeg MIF-a i izračunatu energiju. Nakon toga, sledi formiranje četvorokomponentnih „farmakofora“ od svih mogućih kombinacija MIF-tačaka. Tip MIF-ova koji formiraju četvorokomponentnu „farmakoforu“ kao i rastojanje između njih za sve moguće kombinacije se kodira u vidu farmakofornog fingerprinta. Poređenjem fingerprintova za referentnu i testiranu strukturu vrši se po principu superpozicije svih farmakofornih tačaka sačinjenih od iste kombinacije MIF tačaka. Intenzitet ovog preklapanja se izražava kao skor primenom odgovarajućih funkcija vrednovanja (Slika 3.2) ^{262,265,266}.



Slika 3.2. Princip generisanja i međusobnog poređenja četvorokomponentnih farmakofornih fingerprintova proteina i liganda.

Set poznatih aktivnih i neaktivnih inhibitora SIRT2-a prikupljen je pomoću ChEMBL baze podataka ⁹⁵ koristeći podatke o inhibitornim aktivnostima izražene kroz IC_{50} i Inh% (procenat inhibicije) vrednosti. Molekuli su smatrani aktivnim ukoliko je $IC_{50} \leq 50 \mu M$ i $Inh\% \geq 60\%$. Molekule sa $Inh\% \leq 40\%$ i $IC_{50} \geq 90 \mu M$ su selektovane kao neaktivne. Molekulski fragmenti ($Mw \leq 300$) uklonjeni su iz konačne baze podataka. Molekulski mamci su dobijeni iz maksimalno nepristrasne baze podataka za SIRT2 (eng. *Maximally Unbiased Benchmarking Dataset* – MUDB ²⁶⁷), a ostali su generisani pomoću DUD-E servera (<https://dude.docking.org/>) ²⁶⁸. Za svaku molekulu je izračunata raspodela tautomernih stanja pri fiziološkom pH koristeći ChemAxon cxcalc alat ²⁴², uz zadržavanje tautomera sa predviđenom zastupljenošću većom od 40%. Konačna baza podataka je generisana koristeći FLAP 2.2.1 zadržavajući do 25 minimizovanih konformera za svaki ligand. Konačna baza podataka sadržala je 513 stvarno aktivnih, 114 stvarno neaktivnih molekula i 12239 virtuelnih molekula mamaca. FLAP fingerprintovi (skup četvorokomponentnih farmakoforskih karakteristika) za svaku molekulu izračunati su korišćenjem podrazumevanih podešavanja: 4 osnovne GRID MIF probe (proba H - opisuje oblik džepa za vezivanje; probe N1 i O - opisuje akceptorske i donorske sposobnosti H-veza; DRY proba - procenjuje hidrofobne interakcije) na rezoluciji mreže od 2 Å. Zatim su strukture proteina uvezene u FLAP, a fingerprintovi vezivnog mesta su izračunati na isti način kao i za ligande. Preklapanje između FLAP fingerprintova generisanih za ligande i vezivna mesta kvantifikovano je kroz podrazumevane funkcije vrednovanja GRID MIF proba. Retrospektivni SBVS je sproveden na skupu konformacionih stanja SIRT2:ligand kompleksa dobijenih simulacijama. Za generisanje multistrukturalnog VS modela, izvršena je linearna diskriminantna analiza (LDA) primenom LDA

alata u okviru FLAP programa na celom skupu konformacionih stanja proteina (uključujući i kristalografske strukture)⁶⁸.

Primenom LDA alata generisan je SBVS model kombinacijom maksimalno 3 konformaciona stanja i maksimalno 2 skora u automatskom režimu. Za treniranje LDA modela, 1/3 stvarno aktivnih/neaktivnih molekula su nasumično dodeljeni trening setu, uravnotežavajući odnos aktivnih i neaktivnih molekula. Površina ispod ROC (eng. *Receiver Operating Characteristics* – ROC) krive – ROC_AUC, korišćena je za numeričko kvantifikovanje performansi klasifikacionih VS modela. Osim toga, performanse različitih strukturno-zavisnih VS modela su procijenjene izračunavanjem faktora obogaćenja (eng. *Enrichment factor* – EF) ROC-a (EF1%, EF2% i EF5%), koji kvantifikuju područje pokriveno ROC krivom obogaćenja pri 1%, 2% ili 5% lažno pozitivnih rezultata⁶⁸. Nakon odabira najboljeg SBVS modela, izvršeno je prospektivni SBVS SPECS baze podataka od preko 200,000 komercijalno dostupnih jedinjenja²⁶⁹. Jedinjenja dobijena iz SPECS-a su prefiltrirana korišćenjem „*Rule of Five*“ filtera u Instant JChem 21.2.0 paketu²⁷⁰. Hemijski prostor je projektovan na Kohonenove samoorganizujuće mape izgrađene iz skupa stvarno aktivnih/neaktivnih molekula kao i najbolje rangiranih rezultata skrininga (0.5% ukupno pregledanih jedinjenja - približno 1000 jedinjenja). Mape su izgrađene korišćenjem MACCS RDKit (verzija 2021.03.4)²⁷¹ deskriptora i MiniSom 2.2.9 biblioteke²⁷². Jedinjenja iz manje zastupljenih delova hemijskog prostora su odabrana nasumično i nabavljena od dobavljača za biološku evaluaciju.

3.1.5. Građenje i validacija *in silico* LBVS platforme za otkriće novih SIRT2 inhibitora

Inicijalni podaci o strukturama i biološkim aktivnostima SIRT2 inhibitora su preuzeti sa ChEMBL baze podataka (verzija 30). Prikupljeni su podaci kod kojih je biološka aktivnost bila izražena kao IC₅₀ vrednosti, ili procenat inhibitornog odgovora, dok su hemijske strukture bile izražene u SMILES formatu²⁷³. Dodatna jedinjenja su preuzeta iz patenta US20160376238A1²⁷⁴. Strukture jedinjenja su ekstrahovane iz patenta primenom ChemDataExtractor (v1.3.0) programa²⁷⁵. Sirovi podaci su inicijalno podeljeni na 4 seta podataka (setovi podataka 1-4) na osnovu namenjene primene koja je detaljnije diskutovana dalje u tekstu. Setovi podataka 1-4 su ručno obrađeni uklanjanjem zapisa o bilo kom drugom vidu aktivnosti osim deacetilacione aktivnosti (demiristoilaciona, dedekanoilaciona itd.). Tako obrađeni setovi podataka su dalje podvrgnuti proceduri kanonizacije SMILES zapisa, uklanjanju duplikata, soli i naelektrisanja molekula. Svi koraci pripreme molekula su odrađeni primenom RDKit (v 2021.03.4) Python biblioteke²⁷¹. Duplirana jedinjenja sa višestrukim zapisima o aktivnostima su pre uklanjanja ručno proveravana i zadržavani su samo zapisi čija je vrednost bila najpribližnija srednjoj vrednosti grupe duplikata. Ukoliko je isto jedinjenje posedovalo i IC₅₀, i % inhibicije (Inh%), zadržan je samo zapis of IC₅₀ vrednosti.

U cilju formiranja regresionih modela, IC₅₀ vrednosti sadržane u bazi podataka su konvertovane u pIC₅₀ ($pIC_{50} = -\log_{10}(IC_{50})$), dok su za klasifikacione modele IC₅₀ i Inh% vrednosti enkodirane u određenu klasu. Jedinjenjima je dodeljivana klasa aktivnih SIRT2 inhibitora ukoliko je IC₅₀ ≤ 50 μM, ili Inh% ≥ 80% (testirano pri koncentraciji 200 μM), ili Inh% ≥ 70% (testirano pri koncentraciji 100 μM), ili Inh% ≥ 60% (testirano pri koncentracijama 50 ili 100 μM), ili Inh% ≥ 50% (testirano pri koncentracijama ispod 50 μM). Jedinjenjima je dodeljena klasa neaktivnih SIRT2 inhibitora ukoliko je IC₅₀ ≥ 90 μM, ili Inh% ≤ 40% (testirano pri koncentraciji iznad 100 μM). Ukoliko su jedinjenja istovremeno sadržala i zapis o aktivnosti na izoformama SIRT1 i SIRT3, aktivnosti na ovim izoformama SIRT2 su takođe enkodirane u klase ekvivalentno pravilima navedenim za SIRT2 inhibitoru aktivnost. Pre treniranja modela u okviru svakog od 4 spomenuta seta podataka, jedinjenja su podeljena na trening set (70%) i test set (30%) koristeći algoritam stratifikovane podele podataka na trening set i test set koji je implementiran u scikit-learn (v 1.1.1) Python biblioteku²⁷⁶. Disbalans klasa unutar setova podataka namenjenih građenju klasifikacionih modela

(setovi podataka 2-4) je korigovan preprimenom SMOTE (eng. *Synthetic Minority Oversampling Technique*) algoritma implementiranog unutar *imbalanced-learn* biblioteke (v 0.9.1)²⁷⁷.

3.1.5.a. Računanje molekulskih deskriptora/fingerprintova

Za jedinjenja unutar svakog seta podataka izračunati su MACCS fingerprintovi (166 bitova), ECFP4 i ECFP6 fingerprintovi (1024 bita) i dvodimenzionalni molekularni deskriptori (ukupno 1613 deskriptora). Fingerprintovi su izračunati primenom RDKit Python biblioteke, dok su deskriptori izračunati primenom Mordred v 1.2.0 programa²⁷⁸. Unutar svakog od setova podataka, broj molekulskih deskriptora je dalje redukovano uklanjanjem „NaN“ vrednosti (deskriptori koje nije bilo moguće izračunati). Vrednosti preostalih deskriptora su standardizovane i nakon toga deskriptori koji su imali nisku varijansu su uklonjeni koristeći graničnu vrednost od 0,1. Korelacije između parova deskriptora su izračunati primenom *Pearson*-ovog korelacionog koeficijenta. Ukoliko su dva deskriptora imali korelacioni koeficijent veću od 0,9 zadržan je samo jedan deskriptor. Finalna selekcija deskriptora za modelovanje je izvršena primenom algoritma rekurzivne eliminacije uz ukrštenu validaciju implementiranog u scikit-learn (v 1.1.1) biblioteku²⁷⁶. Klasifikator stabla odlučivanja (eng. *Decision Tree Classifier*) je upotrebljen kao procenitelj uz desetstruku ukrštenu validaciju. Svi navedeni koraci selekcije deskriptora izvedeni su na trening setu svakog od seta podataka.

3.1.5.b. Građenje i evaluacija modela

Za građenje modela mašinskog učenja, upotrebljeno je pet algoritama: algoritam slučajnih šuma (eng. *Random Forests* – RF), algoritam potpornih vektora (eng. *Support Vector Machines* – SVM) za klasifikaciju (eng. *Support Vector Classification* – SVC) i regresiju (eng. *Support Vector Regression* – SVR), algoritam k-najbližih suseda (eng. *K-Nearest Neighbors* – KNN), algoritam ekstremnog gradijentnog pojačanja (eng. *eXtreme Gradient Boosting* – XGBoost) i duboke neuronske mreže (eng. *Deep Neural Networks* – DNN). Regresori i klasifikatori za RF, SVM, XGBoost i KNN modele su izgrađeni primenom scikit-learn biblioteke uz dodatak XGBoost Python biblioteke (v 1.5.1). DNN modeli su trenirani primenom TensorFlow (v 2.9.1) biblioteke²⁷⁹. Hiperparametri svakog modela (osim DNN) su optimizovani uz primenu *Bayes*-ove optimizacije sa petostrukom ukrštenom validacijom kroz scikit-optimize biblioteku (v 0.8.1). Optimizacija hiperparametara i arhitekture mreže DNN modela, izvršena je Python skriptom zasnovanom na funkcionalnosti Keras Tuner-a (v 1.1.1)²⁸⁰. Optimizacija DNN modela takođe je bila zasnovana na *Bayes*-ovu optimizaciju i petostruku ukrštenu validaciju.

Tri različita tipa modela mašinskog učenja su kreirana u zavisnosti od korišćenog seta podataka: regresioni model (set podataka 1), binarni klasifikacioni model (set podataka 2) i višeklasni klasifikacioni modeli (setovi podataka 3 i 4). Interna validacija regresionih modela je izvedena praćenjem koeficijenta determinacije za trening set (R_{int}^2) (jednačina 28) i korelacionog koeficijenta ukrštene validacije (Q^2) (jednačina 29), korena srednje vrednosti kvadrata greške (eng. Root Mean Square Error – RMSE) fitovanja trening seta ($RMSE_{int}$) (jednačina 30) i ukrštene validacije ($RMSE_{CV}$) (jednačina 31):

$$R_{int}^2 = 1 - \frac{\sum_{i=1}^{n_{int}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{int}} (y_i - \bar{y})^2} \quad (28)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (29)$$

$$RMSE_{int} = \sqrt{\frac{\sum_{i=1}^{n_{int}} (y_i - \hat{y}_i)^2}{n_{int}}} \quad (30)$$

$$RMSE_{CV} = \sqrt{\frac{\sum_{n=1}^n (y_i - \hat{y}_{i/i})^2}{n}} \quad (31)$$

gde y_i predstavlja eksperimentalno određene vrednosti zavisne varijable, \hat{y}_i predstavlja predviđene vrednosti, $\hat{y}_{i/i}$ predstavlja odgovor i -tog elementa procenjeno modelom dobijenim bez upotrebe i -tog elementa i n predstavlja broj elemenata⁹⁸.

U okviru interne validacije regresionih modela, primenjena je i metoda Y-randomizacije u sklopu koje je generisano 100 regresionih modela na osnovu nasumično izmešanih podataka koristeći iste vrednosti hiperparametara kao kod optimizovanih modela. Eksterna validacija modela je izvedena upotrebom seta statističkih parametara: koeficijent determinacije test seta (R_{ext}^2) (jednačina 32), RMSE test seta ($RMSE_{ext}$) (jednačina 33)⁹⁸, $Q_{F_n}^2$ metrika (Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2) (jednačine 34-36)^{96,99,100}, r_m^2 metrika (r_m^2 , \bar{r}_m^2 , Δr_m^2) (jednačine 37 i 38)^{101,102} i korelacioni koeficijent podudarnost (eng. concordance correlation coefficient (CCC)) (jednačina 39)¹⁰³:

$$R_{ext}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2} \quad (32)$$

$$RMSE_{ext} = \sqrt{\frac{\sum_{n=1}^{n_{EXT}} (y_i - \hat{y}_i)^2}{n_{EXT}}} \quad (33)$$

$$Q_{F1}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y}_{TR})^2} \quad (34)$$

$$Q_{F2}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y}_{EXT})^2} \quad (35)$$

$$Q_{F3}^2 = 1 - \frac{[\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2] / n_{EXT}}{[\sum_{i=1}^{n_{TR}} (y_i - \bar{y}_{TR})^2] / n_{TR}} \quad (36)$$

$$r_m^2 = r^2 (1 - \sqrt{(r^2 - r_0^2)}) \quad (37)$$

$$\Delta r_m^2 = |r_m^2 - r'^2_m| \quad (38)$$

$$CCC = \frac{2 \sum_{i=1}^{n_{EXT}} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2 + \sum_{i=1}^{n_{EXT}} (\hat{y}_i - \bar{\hat{y}})^2 + n_{EXT} (y_i - \bar{y})^2} \quad (39)$$

U navedenim jednačinama TR predstavlja trening set, EXT predstavlja test set, y_i predstavlja eksperimentalne vrednosti zavisne promenljive, \hat{y}_i predstavlja predviđene vrednosti zavisne promenljive, \bar{y}_i predstavlja srednju vrednost eksperimentalnih vrednosti zavisno promenljivih, $\bar{\hat{y}}$ predstavlja srednju vrednost predviđenih vrednosti zavisno promenljivih. r_0^2 i r^2 su koeficijenti determinacije regresione prave izračunate upotrebom eksperimentalnih i predviđenih vrednosti test seta sa forsiranjem regresije kroz nulu, (r_0^2) ili bez (r^2). r_m^2 je izračunata upotrebom eksperimentalnih vrednosti na y-osi dok je r'^2_m izračunat koristeći iste vrednosti na x-osi. \bar{r}_m^2 je srednja vrednost r_m^2 i r'^2_m .

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (40)$$

$$h^* = \frac{3(m+1)}{p} \quad (41)$$

Domen primenjivosti za regresione modele je računat primenom *leverage* metode²⁸¹. *Leverage* vrednosti (h_i) su izračunate na osnovu jednačine (40) gde je X matriks najznačajnijih deskriptora/bitova molekula trening seta, dok je x_i vektor deskriptora/bitova testiranog jedinjenja. Granična *leverage* vrednost (h^*), je izračunata primenom jednačine (41), gde m predstavlja broj

deskriptora/bitova, dok je p broj molekula trening seta. Značajnost deskriptora/bita unutar svakog modela računata je primenom pristupa permutacione značajnosti implementiranog u scikit-learn biblioteku koristeći 30 ponavljanja.

Klasifikacioni modeli su evaluirani koristećenjem: balansirane tačnosti (eng. *Balanced Accuracy* – BA), opoziva (eng. *Recall*), preciznosti, F1-skora, *Matthews*-ovog korelacionog koeficijenta (eng. *Matthew's correlation coefficient* – MCC) i površine ispod ROC (eng. *Receiver Operating Characteristics* – ROC) krive (ROC_AUC). Celokupna metrika je izvedena iz konfuzione matrice kreirane praćenjem broja pravih pozitivnih predviđanja (eng. *True Positive* – TP), pravih negativnih predviđanja (eng. *True Negative* – TN), lažno pozitivnih predviđanja (eng. *False Positive* – FP) i lažno negativnih predviđanja (eng. *False Negative* – FN). Opoziv (senzitivnost predviđanja) je računat prema formuli $Opoziv = TP / (TP + FN)$ i predstavlja sposobnost modela da ispravno klasifikuje aktivno jedinjenje kao aktivno računajući sva pozitivna predviđanja. Specifičnost predviđanja je računata prema formuli $Specifičnost = TN / (TN + FP)$ i predstavlja sposobnost modela da ispravno klasifikuje negativno jedinjenje kao negativno uzimajući u obzir sva negativna predviđanja. Balansirana tačnost predstavlja srednju vrednost senzitivnosti i specifičnosti. Preciznost reflektuje sposobnost klasifikacionog modela da ispravno klasifikuje sva aktivna jedinjenja kao pozitivno predviđanje ($Preciznost = TP / (TP + FP)$), dok je F1-skor računat kao harmonizovana srednja vrednost preciznosti i opoziva ($F1 = 2 \times (Preciznost \times Opoziv) / (Preciznost + Opoziv)$) što čini ovaj skor direktno povezanim sa robusnošću i preciznošću klasifikatora. MCC predstavlja kompleksnu metriku kojoj se sumiraju sve kategorije konfuzione matrice (jednačina 42)²⁸².

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (42)$$

Vrednosti MCC iznad 0 ukazuju na dobre karakteristike klasifikatora u sve 4 kategorije konfuzionih matrica. ROC kriva je kreirana praćenjem opoziva u odnosu na specifičnosti pri različitim frakcijama seta podataka. Stoga, površina ispod ROC krive (ROC_AUC) se može dovesti u vezu sa verovatnoćom da klasifikator rangira nasumično odabrana aktivna jedinjenja u sortiranoj listi iznad nasumično odabranih neaktivnih jedinjenja. U slučaju višeklasnih klasifikacionih modela, vrednosti ROC_AUC, Preciznosti, Opoziva i F1-skora su računate kao makro srednje vrednosti primenom jedan-prema-ostalima pristupa. Dodatna evaluacija klasifikacionih modela vršena je pomoću seta podataka virtuelnih molekula mamaca. Virtuelni molekuli mamci su kreirani pomoću DUD-E servera i nadalje su priključeni test setu²⁶⁸. Za grafičku interpretaciju atomskih doprinosa binarnog i višeklasnih modela i kreiranje hipoteza odnosa strukture i aktivnosti, upotrebljene su mape sličnosti (eng. *Similarity Maps*) na osnovu RDKit implementacije algoritma razvijenog od strane *Riniker* i saradnika²⁸³. Analiza pokrivenosti hemijskog prostora izvedena je primenom Kohonenovih samoorganizujućih mapa izgrađenih korišćenjem MACCS RDKit (verzija 2021.03.4) fingerprintova i MiniSom 2.2.9 biblioteke kako je objašnjeno u sekciji 3.1.4. Prospektivni skrining je izveden na maloj biblioteci jedinjenja dostupnih u trenutku izvođenja studije. Jedinjenja selektovana prospektivnim skriningom su testirana na procenat inhibicije prema protokolu opisanom u poglavlju 3.1.7.

3.1.6. Sinteze novih selektivnih SIRT2 inhibitora

- **Sintetski prekursori, reagensi i rastvarači**

1-Hidroksibenzotriazol hidrat (HOBt), 98%, Acros Organics

3-nitrobenzil bromid, ≥98%, Acros Organics

3-Fenilpropanska kiselina, 99%, Acros Organics

3-[4-(Trifluorometil)fenil]propanska kiselina, 95%, Acros Organics

4-Piridin-4-ilmetil-anilin, 97%, Thermo Scientific

4-Dimetilaminopiridin (DMAP), 99%, Acros Organics
Aceton, HPLC čistoće, Sigma Aldrich
Acetonitril, HPLC čistoće, Sigma Aldrich
Celite® 545, Sigma Aldrich
Di-tercbutil dikarbonat, 99%, Acros Organics
Dihlormetan (DCM), HPLC čistoće, Sigma Aldrich
Etil-acetat, p.a, Fisher Chemical
Fenilsirćetna kiselina, 98,5%, Acros Organics
Izonikotinska kiselina, 99%, Sigma Aldrich
N-(3-Dimetilaminopropil)-*N'*-etilkarbodiimid hidrohlorid (EDCI), ≥98%, Sigma Aldrich
N,N-Diizopropiletilamin (DIPEA), 99+%, Acros Organics
Natrijum-nitrit, 99%, Acros Organics
Natrijum-azid, 99%, Acros Organics
Natrijum-borohidrid, >98%, Acros Organics
Nikl(II)-hlorid heksahidrat, 99,999%, Acros Organics
N,N-Dimetilformamid (DMF), ≥99.8 %, Fisher Chemical
p-aminofenol i *m*-aminofenol, ≥97%, Acros Organics
trans-Cimetna kiselina, 98+%, Acros Organics
terc-Butanol (*t*-BuOH), 99,5%, Acros Organics
Trifluorsirćetna kiselina (TFA), HPLC čistoće, Sigma Aldrich
Silikagel 60 za hromatografiju na koloni (veličina čestica 0,035-0,075 mm, veličina pore 60 Å) (Sigma Aldrich, Steinheim, Nemačka)
Rastvarači za NMR eksperimente: Deutero hloroform (CDCl₃) i deutero dimetilsulfoksid (DMSO-*d*₆)

- **Oprema**

Analitička vaga, Adventurer-Pro (proizvođač Ohaus Corporation)
Magnetna mešalica sa grejačem (proizvođač Tehtnica Rotamix SHP-10)
Ploče za tankoslojnu hromatografiju, dimenzije 20 x 20 cm, Kieselgel 60 F254, (proizvođač Merck)
Komora sa izvorom ultraljubičastog zračenja, UV-Cabinet (Camag, Muttenz, Švajcarska)
Rotacioni vakuum uparivač (Heidolph Rotary Evaporator)
Rotacioni vakuum uparivač visokih performansi (Rotary vane pump RZ 2.5, Vacuumbrand, Wertheim, Nemačka)
Vakuum sušnica (Thermo Heraeus, Fisher Scientific, UK, VT 6025)

- **Procedure**

Generalna procedura 1. Sinteza *N*-Boc derivata aminofenola. Komercijalno dostupni *p*-aminofenol (1 ekvivalent), ili *m*-aminofenol (1 ekvivalent) su rastvoreni uz mešanje u tetrahidrofuranu (THF), nakon čega je dodat di-*tert*-butil dikarbonat (1 ekvivalent). Reakciona smeša je mešana 24h na sobnoj temperaturi, zaštićena od svetlosti. Tok reakcije je praćen TLC hromatografijom uz mobilnu fazu etil acetat : petrol etar = 1:1. Nakon završetka reakcije, reakciona smeša je koncentrovana na vakuum uparivaču, a ostatak prenet u levak za ekstrakciju i raspodeljen između vode i etil-acetata (2x). Organski sloj je ispran zasićenim rastvorom natrijum-hlorida i sušen anhidrovanim magnezijum-sulfatom i nakon filtracije organski rastvarač je uparen na vakuum uparivaču.

Generalna procedura 2. Nukleofilna supstitucija u prisustvu K_2CO_3 . *N*-Boc aminofenoli ND3_1, ili ND4_1 (1 ekvivalent) su rastvoreni uz mešanje u acetonitrilu, nakon čega je dodat kalijum karbonat (1,2 ekvivalenta). Smeša je mešana 30 minuta, nakon čega je dodat 3-nitrobenzil bromid (1 ekvivalent). Smeša je mešana 24h na sobnoj temperaturi. Tok reakcije je praćen TLC hromatografijom uz mobilnu fazu dietil etar : petrol etar = 1:1. Nakon završetka reakcije, reakciona smeša je koncentrovana na vakuum uparivaču, a ostatak prenet u levak za ekstrakciju i raspodeljen između vode i etil-acetata (2x). Organski sloj je ispran zasićenim rastvorom natrijum-hlorida, sušen anhidrovanim magnezijum sulfatom i filtriran. Nakon filtriranja organski rastvarač je uparen na vakuum uparivaču. Dobijeni čvrst proizvod je ispran malom zapreminom petrol etra i osušen.

Generalna procedura 3. Redukcija nitro grupe natrijum-borohidridom. Nitro jedinjenje (ND3_2, ND4_2, ili ND24_C) (1 ekvivalent) je rastvoreno u smeši acetonitril:voda (3:0,5). U smešu je dodat nikel(II)-hlorid heksahidrat (0,2 ekvivalenta) i u narednih 5 min u porcijama je dodavan natrijum-borohidrid (4 ekvivalenta). Smeša je mešana još 1h na sobnoj temperaturi. Nakon završetka reakcije, u balon je dodata destilovana voda i smeša je filtrirana kroz celit (Celite® 545). Filtrat je prenet u levak za ekstrakciju i proizvod reakcije je ekstrahovan etil-acetatom. Organski sloj je ispran zasićenim rastvorom natrijum-hlorida, sušen anhidrovanim magnezijum-sulfatom i filtriran. Nakon filtriranja, organski rastvarač je uparen na vakuum uparivaču.

Generalna procedura 4. Kuplovanje amida u sintezi jedinjenja ND3_4, ND4_4, ND5_4, ND6_4 i ND7_4. Anilinski derivat (ND3_2, ili ND4_2) (1 ekvivalent) je rastvoren uz mešanje u acetonitrilu. U reakcionu smešu je dodata odgovarajuća karboksilna kiselina (1,2 ekvivalent), EDCI (1 ekvivalent), HOBt (0,1 ekvivalent), DMAP (1 ekvivalent) i DIPEA (5 ekvivalenta). Smeša je mešana na sobnoj temperaturi 24h. Nakon završetka reakcije, smeša je koncentrovana na vakuum uparivaču, dodat je zasićeni rastvor $NaHCO_3$ i proizvod reakcije ekstrahovan etil-acetatom. Organski sloj je dodatno ispran vodom i zasićenim rastvorom natrijum-hlorida, nakon čega je sušen anhidrovanim magnezijum-sulfatom i filtriran. Nakon uklanjanja organskih rastvarača, proizvod reakcije je prečišćen hromatografijom na koloni (mobilna faza: petrol-etar : dihlormetan : etil-acetat : glacijalna sirćetna kiselina = 1:0,5:0,45:0,05). Frakcija sa supstancom je dodatno isprana zasićenim rastvorom $NaHCO_3$, a vodeni sloj je dodatno ekstrahovan etil-acetatom. Spojeni organski slojevi su sušeni anhidrovanim magnezijum-sulfatom i filtrirani. Nakon filtriranja, organski rastvarač je uparen na vakuum uparivaču.

Generalna procedura 5. Uklanjanje *N*-Boc zaštite. *N*-Boc jedinjenje (ND3_4, ND4_4, ND5_4, ND6_4, ili ND7_4) (1 ekvivalent) je rastvoreno u 25% rastvoru trifluor-sirćetne kiseline u dihlormetanu i reakciona smeša je mešana na sobnoj temperaturi. Reakcija je praćena TLC hromatografijom uz mobilnu fazu dihlormetan : etil-acetat = 1:1. Nakon završetka reakcije, u balon je dodat $NaHCO_3$ (do neutralne reakcije). Vodeni rastvor je prebačen u levak za ekstrakciju i proizvod ekstrahovan etil-acetatom (3x). Organski sloj je ispran zasićenim rastvorom natrijum-hlorida, sušen anhidrovanim magnezijum-sulfatom, filtriran i uparen na vakuum uparivaču.

Proizvod je dalje prečišćen hromatografijom na koloni (mobilna faza dihlormetan : etil-acetat = 1:1).

Generalna procedura 6. Građenje azida. Anilinski derivati (ND3_5, ND4_5, ND5_5, ND6_5, ili ND7_5) (1 ekvivalenat), rastvoreni u smeši glacijalne sirćetne kiseline i vode u odnosu 5:2 (c = 0,3 M), su mešani na ledenom kupatilu. Zatim je u rastvor dodat 1,3 M vodeni rastvor natrijum-nitrita (1,3 ekvivalenata). Reakciona smeša je mešana 30 minuta na istoj temperaturi. Nakon toga je u porcijama dodat 1,5 M vodeni rastvor natrijum-azida (1,5 ekvivalenata). Reakciona smeša je mešana još 2 sata pod istim uslovima. Nakon završetka reakcije, vodeni rastvor je neutralizovan zasićenim rastvorom NaHCO₃ i proizvod reakcije je ekstrahovan etil-acetatom (3x). Organski sloj je ispran zasićenim rastvorom natrijum-hlorida, sušen anhidrovanim magnezijum-sulfatom i filtriran, nakon čega je organski rastvarač je uparen na vakuum uparivaču. Proizvod reakcije je prečišćen hromatografijom na koloni (mobilna faza etil-acetat:dihlormetan = 1:1).

Sinteza 4-prop-2-inil-morfolina. Morfolin (2,5 ekvivalenata) je rastvoren u dihlormetanu uz mešanje i u smešu je u porcijama dodat propargil-bromid (1 ekvivalenat). Reakciona smeša je refluktovana 4h. Nakon završetka reakcije, reakciona smeša je razblažena vodom, supstanca je ekstrahovana dihlormetanom (3x). Organski sloj je ispran zasićenim rastvorom natrijum-hlorida, sušen anhidrovanim magnezijum-sulfatom, filtriran i rastvarač je uparen na vakuum uparivaču. Proizvod reakcije je prečišćen hromatografijom na koloni (mobilna faza – aceton).

Generalna procedura 7: Bakar(I) katalizovana azid-alkin cikloadicija. Azid (ND3_6, ND4_6, ND5_6, ND6_6, ili ND7_6) (1 ekvivalenat), 4-prop-2-inil-morfolin (1 ekvivalenat) i Na-askorbat (3 ekvivalenata) su rastvoreni u smeši terc-butanol : dihlormetan : voda (1:0,5:0,5) uz mešanje nakon čega je dodat bakar(II)-sulfat pentahidrat (2 ekvivalenata). Reakcija je praćena TLC hromatografijom uz mobilnu fazu etil-acetat:aceton =1:1. Nakon završetka reakcije, u reakcionu smešu je dodat 14% rastvor amonijaka, smeša je prenetu u levak za ekstrakciju i proizvod reakcije je ekstrahovan etil-acetatom (3x). Organski sloj je ispran zasićenim rastvorom natrijum-hlorida, sušen anhidrovanim magnezijum-sulfatom, filtriran i organski rastvarač je uparen na vakuum uparivaču. Proizvod reakcije je prečišćen primenom hromatografije na koloni (mobilna faza etil-acetat:aceton).

Generalna procedura 8: Kuplovanja amida u sintezi jedinjenja ND18, ND19, ND20, ND23, ND24_A i ND24_B. Odgovarajući anilinski derivat (benzen-1,3-diamin, 4-piridin-4-ilmetil-anilin, ili ND24_A) (0,9 ekvivalenata) je rastvoren uz mešanje u smeši dihlormetana i dimetilformamida (10:0,5). U reakcionu smešu je dodata odgovarajuća karboksilna kiselina (1 ekvivalenat), EDCI (1 ekvivalenat), HOBt (1 ekvivalenat) i trietilamin (3 ekvivalenata). Reakciona smeša je mešana na sobnoj temperaturi 24h. Nakon završetka reakcije, smeša je koncentrovana na vakuum uparivaču, dodat je zasićeni rastvor NaHCO₃ i proizvod reakcije ekstrahovan etil-acetatom. Organski sloj je dodatno ispran destilovanom vodom i zasićenim rastvorom natrijum-hlorida, nakon čega je sušen anhidrovanim magnezijum-sulfatom i filtriran. Nakon uklanjanja organskih rastvarača, proizvod reakcije je prečišćen preparativnom TLC hromatografijom (mobilna faza: dihlormetan:aceton).

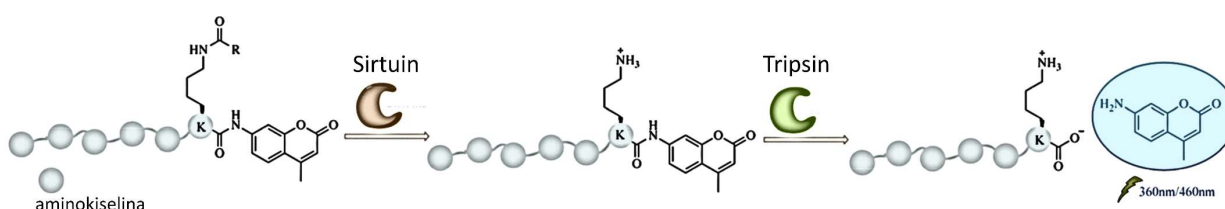
Generalna procedura 9: Reduktivna aminacija: Anilinski derivat ND24_A (1,1 ekvivalenta) je rastvoren u dihlormetanu uz mešanje i dodati su 4-nitrobenzalhid (1 ekvivalenat), *Hantzsch*-ov estar (1,4 ekvivalenata), trifluorsirćetna kiselina (0,5 ekvivalenata) i anhidrovani magnezijum-sulfat (4 ekvivalenata). Reakciona smeša je refluktovana 12 časova uz prisustvo aktiviranih molekulskih sita. Nakon završetka reakcije, sita su uklonjena filtriranjem, a filtrat je koncentrovana na vakuum uparivaču, dodat je zasićeni rastvor NaHCO₃ i proizvod reakcije ekstrahovan etil-acetatom. Organski sloj je ispran destilovanom vodom i zasićenim rastvorom natrijum-hlorida, nakon čega je sušen anhidrovanim magnezijum-sulfatom i filtriran. Nakon uklanjanja organskih rastvarača, proizvod reakcije je prečišćen preparativnom TLC hromatografijom (mobilna faza: dietil-etar:petrol-etar 1:1).

Prekursori i konačni proizvodi su okarakterisani NMR ($^1\text{H}/^{13}\text{C}$) spektroskopijom.

3.1.7. Biološka testiranja

3.1.7.a. *In vitro* enzimski testovi za SIRT1, SIRT2 i SIRT6

Deacetilaciona aktivnost sirtuina je praćena primenom eseja zasnovanih na principu rada *Fluor de Lys* testa. Za praćenje deacetilacione aktivnosti sirtuina *Fluor de Lys* koristi fluorogeni supstrat izveden iz p53 (SIRT1-2) ili H3 (SIRT6) proteina na čijem C-terminalnom kraju je kovalentno vezan 7-amino-4-metilkumarinom (AMC). Fluorescencija AMC je ugašena kroz konjugaciju sa C-terminalnim krajevima kratkog acetil-lizin peptida. Ovakav fluorogeni supstrat nije supstrat za hidrolizu C-terminalnog kraja tripsinom. Nakon deacetilacije sa sirtuinima, usled oslobađanja ϵ -amino grupe lizina, peptid postaje supstrat za tripsin. Hidroliza tripsinom oslobađa fluoroforu što dovodi do pojave fluorescencije (Slika 3.3)²⁸⁴.



Slika 3.3. Princip SIRT1-2 eseja sa fluorogenim supstratom (prilagođeno na osnovu reference²⁸⁵)

- **Protokol za *in vitro* enzimski test za SIRT1-2**

Deacetilacione aktivnosti SIRT1-2 u prisustvu testiranih jedinjenja su određene kao što je prethodno opisano za fluorimetrijske testove *Fluor de Lys* testom (Enzo Life Sciences, Ann Arbor, MI, SAD)²⁸⁶. Reakcija je započeta inkubacijom enzima (SIRT1 ili SIRT2) sa smešom reakcije koja sadrži acetilovan peptidni supstrat BioMol KI177 za SIRT1 i KI179 za SIRT2 (0,7 Km: 58 μM za SIRT1, 198 μM za SIRT2), NAD^+ (0,9 Km: 558 μM za SIRT1, 547 μM za SIRT2) i 5% dimetil sulfoksid (DMSO) ili jedinjenja rastvorena u DMSO u koncentraciji od 5%. Inkubacija je izvršena na 37 °C, tokom 1 sata. Dodati su razvijajući i nikotinamid (2 mM u puferu za histon-deacetilazu (HDAC), dajući ukupni volumen od 50 μl), a inkubacija je nastavljena tokom 45 min na 37 °C. Fluorescencija je izmerena pomoću fluorimetrijskog čitača EnVision 2104 (PerkinElmer, Waltham, MA, SAD) sa talasnim dužinama ekscitacije i emisije 370 nm i 460 nm. Svi eksperimenti su izvedeni u triplikatu. Jedinjenja su prvo ispitana u pogledu inhibicione aktivnosti SIRT1-2 pri koncentraciji od 200 μM , ili 100 μM radi dobijanja procenta inhibicije. Za jedinjenja ND18 i ND85 ispitano je i dozno-zavisni odgovor pri koncentracijama u opsegu od 1 μM do 1000 μM radi dobijanja IC_{50} vrednost. Procenat SIRT2 aktivnosti je prilagođen sigmoidnoj dozno-zavisnoj jednačini, a zatim su IC_{50} vrednosti izračunate pomoću softvera Graph Pad Prism verzija 6 (Kalifornija, SAD).

- **Protokol za *in vitro* enzimski test za SIRT6**

Inhibitorna aktivnost ispitivanih jedinjenja na SIRT6 je praćena upotrebom fluorogenog deacetilacionog testa prethodno opisanog u literaturi^{286,287}. Ukratko, DMSO kontrola (konačna koncentracija 3%) ili supstance (200 μM) u DMSO-u (konačna koncentracija rastvarača 3%) inkubirane su sa GST-SIRT6 (4,5 μg /bazenu), AMC obeleženim supstratom (Ac-RYQK(Ac)-AMC (CASLO ApS (Danska): 1,28 Km: 320 μM)), i NAD^+ (1,28 Km: 960 μM) u Tris puferu [25 mM, pH 8,0] na +37 °C tokom 90 min. Reakcije su prekinute dodavanjem 50 μl rastvora za razvijanje (6 $\mu\text{g}/\mu\text{l}$ tripsin i 40 mM nikotinamida) u svaki bazen. Reakcija je inkubirana 30 min na sobnoj temperaturi i fluorescencija je merena sa talasnim dužinama ekscitacije i emisije od 380 i 440 nm, redom, korišćenjem EnVision 2104 Multilabel čitača (PerkinElmer, Finska). Sva ispitivanja su izvršena u triplikatu.

3.1.7.b. Protokol za ispitivanje uticaja novootkrivenih jedinjenja na vijabilnost tumorskih ćelija adenokarcinoma dojke

Trostruko negativne (ER-, PR- i HER2-) ćelije adenokarcinoma dojke MDA-MB-231 (ATCC® HTB-26™) su gajene u DMEM:Ham's F-12 (1:1) medijumu (Sigma-Aldrich, Merck KGaA, Darmstadt, Nemačka) suplementiranoj sa penicilinom, streptomycinom i 10% fetalnim telećim serumom u atmosferi od 5% CO₂ na 37°C u inkubatoru uz atmosferu povećane vlažnosti. Procenat preživljavanja MDA-MB-231 ćelija nakon tretmana ND18 i ND19 je procenjen MTT testom, kako je opisano u literaturi²⁸⁸. Ukratko, oko 4x10³ MDA-MB-231 ćelija je zasejano u mikrotitarske ploče sa 96-bazena i tretirano sa ND18 jedinjenjem u šest različitih koncentracija (80, 40, 20, 10, 5 i 2,5 µM), pri čemu je svaka koncentracija testirana u triplikatu. Mikrotitarske ploče su inkubirane 72 sata. Na kraju inkubacionog vremena, u bazene dodato je 20 µL MTT rastvora (3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolijum bromid) (5 mg/mL) (Sigma-Aldrich, Merck KGaA, Darmstadt, Nemačka) i inkubacija je nastavljena još 4 sata. Natrijum dodecil sulfat (SDS) (Merck KGaA, Darmstadt, Nemačka) (10%) je zatim dodat (100 µL) u svaki bazen i inkubacija je nastavljena na 37°C preko noći. Dodatkom SDS-a zaustavljena je hemijska konverzija MTT-a u formazan. Procenat preživljavanja ćelija je izračunat nakon metenja apsorbance na 570nm pomoću čitača za mikrotitarske ploče koristeći matematičku relaciju 43:

$$\% \text{ ćelijske vijabilnosti} = 100 * \frac{A_t - A_b}{A_c - A_b} \quad (43)$$

gde A_t predstavlja srednju vrednost apsorbance tretiranih ćelija, A_b srednju vrednost apsorbance koja potiče od rastvorenog jedinjenja u podlozi i A_c srednju vrednost apsorbance koja potiče od netretiranih ćelija. Citotoksična koncentracija pri kojoj dolazi do redukcije vijabilnosti ćelija za 50% u odnosu na kontrolnu (netretirane ćelije) jeste IC₅₀, odnosno GI₅₀.

3.1.7.c. Analiza raspodele ćelija po fazama ćelijskog ciklusa

MDA-MB-231 ćelije (2x10⁵ ćelija/bazenu) su zasejane u ploče sa 6 bazena i tretirane sa koncentracijom od 20 µM jedinjenja ND18 i ND19 u triplikatima. Nakon inkubacije (72h), ćelije su skupljene i fiksirane hladnim etanolom. Fiksirane ćelije su isprane sa PBS-om, a zatim inkubirane sa RNase A (1 mg/mL) (Sigma-Aldrich, Merck KGaA, Darmstadt, Nemačka) tokom 30 minuta na 37 °C. Neposredno pre analize ćelija na protočnom citometru, ćelije su obojene propidijum jodidom (PI) (Sigma-Aldrich, Merck KGaA, Darmstadt, Nemačka) u koncentraciji od 400 µg/mL. Intenzitet fluorescencije kompleksa PI-DNA u ćelijskoj suspenziji meren je pomoću FACS Calibur Becton Dickinson protočnog citometra (Becton Dickinson, Heidelberg, Nemačka). Rezultati su prikazani kao procenat ćelija u svim fazama ćelijskog ciklusa.

3.1.7.d. Merenje apoptoze protočnom citometrijom

Apoptotska ćelijska smrt je proučavana korišćenjem metode dvokolornog bojenja- Aneksin V-fluorescein izocijanat sa 7-amino-aktinomycinom D (7-AAD) (BD Pharmingen, San Diego, CA, SAD), gde su uzorci pripremljeni prema uputstvima proizvođača. Analiza protočnom citometrijom je izvršena korišćenjem citometra FACS-Calibur uz korišćenje softvera Cell Quest (Becton Dickinson, Heidelberg, Nemačka).

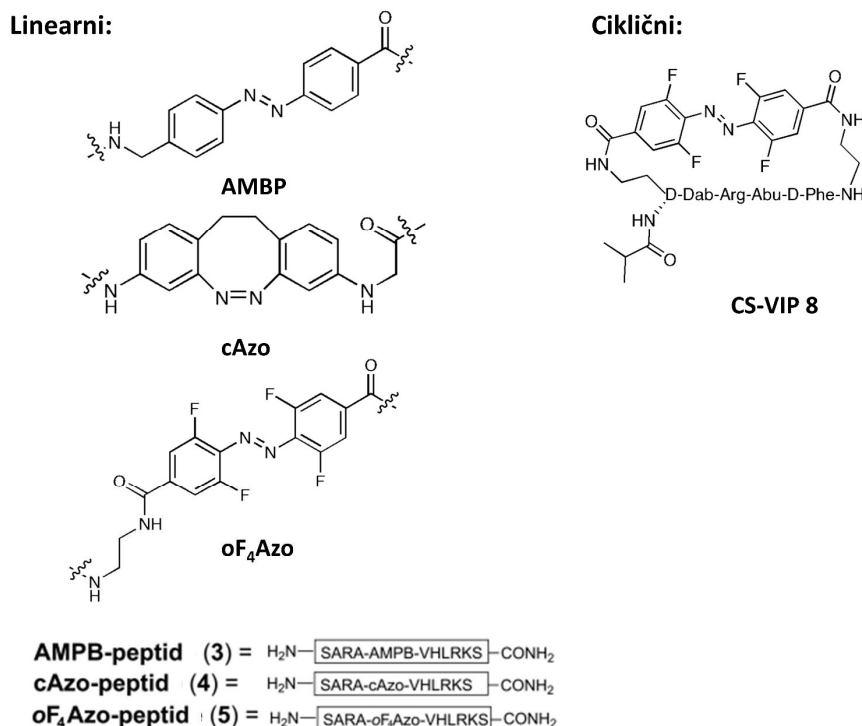
3.1.7.e. Analiza ekspresije E-kadherina protočnom citometrijom

MDA-MB-231 ćelije su tripsinizovane i isprane PBS-om. Zatim su resuspendovane u PBS-u sa 0,2% govedeg serumskog albumina (BSA) nakon čega je dodato primarno zečji anti-E-kadherin (1:100 razblaživanje, Cell Signaling Technology (CST), Danvers, Massachusetts, SAD) antitelo. Inkubirane su 1 h na ledu. Ćelije su jednom isprane u PBS-om, resuspendovane u PBS-u koji sadrži odgovarajuće sekundarno antitelo konjugovano sa fluorescein izocijanatom (razblaženje 1:100; BD Pharmingen, San Diego, CA, SAD), i inkubirane 30 minuta na ledu. Ćelije su jednom isprane u

PBS-u i fiksirane 1% formaldehida. Fluorescencija ćelija je analizirana na protočnom citometru korišćenjem softvera Cell Quest (Becton Dickinson, Heidelberg, Nemačka).

3.2. Molekulsko modelovanje grupe fotoreaktivnih WDR5 inhibitora

Za ispitivanje konformacione dinamike protein-ligand interakcija WDR5 u kompleksu sa fotoreaktivnim inhibitorima, 4 poznata azobenzenska inhibitora (3 linearna i 1 ciklični) sa profilima bioloških aktivnosti su preuzeti iz literature (Slika 3.4). Sva 4 inhibitora su ispitivana u po 2 različita fotostacionarna stanja koja odgovaraju *cis* i *trans* izomerima azobenzenske strukture.^{213,214}



Slika 3.4. Strukture *trans* izomera WDR5 inhibitora korišćenih u istraživanjima predstavljenim u okviru disertacije. U istraživanjima su ispitivani i *cis* izomeri. D-Dab – D-diaminobuterna kiselina, Abu – aminobuterna kiselina.

3.2.1. Modelovanje protein-ligand interakcije linearnih fotoreaktivnih peptidomimetskih inhibitora WDR5

3.2.1.a. Molekulski doking

Za molekulski doking liganda u MLL vezivno mesto WDR5 korišćen je softver GOLD 5.6²⁸⁹. Doking peptida kao inhibitora PPI i dalje predstavlja izazov zbog značajne fleksibilnosti peptida, velikog broja stepeni konformacione slobode kao i teškoća u identifikaciji vezivnih regiona na površini proteina²⁹⁰. Iako postoje specijalizovani programi za doking peptidnih inhibitora PPI, GOLD algoritam, koji je prvobitno namenjen za doking malih molekula, pokazao se kao efikasna alternativa, posebno u slučajevima kada postoji dovoljno eksperimentalnih podataka o načinu interakcije²⁹¹. Primena specijalizovanih programa za doking peptida i proteina otežana je činjenicom da je fotoreaktivna azobenzenska grupa u ligandima uvedena u sredinu sekvence peptida, čineći ih nepodesnim protokolima poznatih specijalizovanih programa. Korišćenjem poznate konformacije SARA-sekvence dobijene u oviru kristalografski razrešenog kompleksa WDR5: AMPB-peptidnog inhibitora (PDB ID: 5M23), postavljen je postupak dokinga u dva koraka po principu fragmentnog dokinga:

1) Doking fragmenta SARA-Azo (gde Azo označava azobenzenski/diazocinski fragment) je izvršen u strukturu prethodno spomenutog kompleksa PDB ID: 5M23 korišćenjem substrukturnih pozicionih ograničenja na SARA-sekvencu. Konformacija SARA-sekvence je zadržana iz PDB-a, dok je Azo ručno dodat u SARA-sekvencu. Konformacija ručno generisanih liganada je relaksirana i dodatno optimizovana kroz 20 ns MD simulacije ovog liganda izvedenoj uz poziciona ograničenjima na SARA-sekvencu korišćenjem Gromacs 5.1.4 softvera (detaljnije objašnjeno u odeljku posvećenom molekulske dinamičkim simulacijama). Usled velikog broja konformacionih stepeni slobode, ovakav vid ograničenog uzorkovanja konformacionih stanja peptidnog fragmenta omogućio je efikasnije rešavanje problema algoritma pretrage molekuskog dokinga;

2) Doking celog molekula izvršen je ograničavanjem SARA-Az položaja dobijenog u prethodnom koraku dokinga. Peptidni fragment (-VHLRKS sekvence) je izgrađen u DS Visualizeru u β -strukturi. Konačna konformacija liganda VHLRKS sekvence generisana je nakon 20 ns MD simulacije liganda sa dodatnim ograničenjima položaja na SARA-sekvencu, koristeći Gromacs 5.1.4 softver (detaljnije objašnjeno u odeljku 3.2.1.b).

Za molekulske doking korišćena je PLP scoring funkcija, sa 50 generisanih poza i maksimalnom fleksibilnošću računatom za ligande. Mesto vezivanja je definisano kao područje od 8 Å oko kristalnog liganda. Težinski faktori finkcije ograničenja su podešeni na 20 za SARA-sekvencu i 5 za azobenzenski fragment. Samo predviđene poze koje su pratile bioaktivnu konformaciju SARA-sekvence iz PDB ID: 5M23 su zadržane za MD simulacije.

3.2.1.b. Simulacije molekulske dinamike

MD protokol uključivao je definisanje azobenzenskih jedinjenja kao novih rezidua. Parametri za azobenzene dobijeni su putem ParamChem veb servera²³⁷ i ručno su uključeni u Charmm36 polje sila²⁴⁴. Za sve MD simulacije korišćen je GROMACS 5.1.4 softver. Protonacija kompleksa proteina i liganda određena je pomoću PROPKA servera²⁴⁰. Protein i aminokiselinske rezidue liganada su parametrizovane CHARMM36 poljem sile. MD simulacija je izvedena u oktaedarskoj simulacionoj kutiji uz primenu TIP3P modela vode. Sistem je preliminarno minimiziran sa 5000 koraka metodom gradijenta strmog spusta. Sistem je zatim zagrejan na 310 K tokom 250 ps u NVT ansamblu sa vremenskim korakom integracije jednačine kretanja od 1 fs (leapfrog algoritam). Nakon toga, pritisak je izjednačen na 1 atm tokom 500 ps NPT simulacije sa vremenskim korakom od 2 fs (leapfrog). U koracima ekvibracije, postavljena su harmonijska poziciona ograničenja na osnovni skelet proteina sa konstantom od 1000 kJ /mol·Å². Poziciona ograničenja korišćena u ekvibracionim koracima postepeno su uklonjena tokom 1 ns simulacije sa vremenskim korakom od 2 fs. Proizvodne simulacije (20 ns) izvršene su u NPT ansamblu na 310 K bez faktora ograničenja, osim u simulacijama pripreme liganada za doking gde su poziciona ograničenja ostavljena na -SARA- sekvenci. Primenjeni su Nose-Hoover termostat, Parrinello-Rahman barostat, LINCS algoritam za ograničenja (sve veze sa H atomima) i *particle mesh Ewald* algoritam za računanje elektrostatskih interakcija dugog dometa. Nakon završetka simulacije, Gromos proces klasterovanja²⁵⁰ trajektorije upotrebljen je za dobijanje odgovarajućih reprezentativnih struktura. MD simulacije su upotrebljene prevashodno za validaciju i usavršavanje kompleksa generisanih molekularnim dokingom. Veći broj ponavljanja izveden je za svaki ligand. Za dalju MM/PBSA analizu zadržane su samo simulacije u kojima je ARA sekvencu bila stabilna u aktivnom mestu. MM/PBSA metoda, kao metoda proračuna slobodne energije krajnjeg stanja, dokazala se kao korisno sredstvo za dodatnu analizu rezultata simulacija²⁹². Iako ova metoda nije dovoljno precizna za predviđanje apsolutnih vezivnih slobodnih energija, ima veliku primenu u racionalizaciji eksperimentalnih rezultata (vezivnih afiniteta) serije sličnih liganda i predviđanju afiniteta novih liganada²⁹²⁻²⁹⁴. Poslednjih 7 ns MD trajektorija je upotrebljeno za MM/PBSA proračune primenom GMXPBSA 2.1. Za proračun je upotrebljen svaki 18. segment trajektorije. Dielektrična konstanta rastvora je podešena na 1, dok su svi drugi parametri zadržani na podrazumevanim vrednostima. Entropijski doprinos je zanemaren u proračunima.

3.2.2. Modelovanje protein-ligand interakcije cikličnog fotoreaktivnog peptidomimetskog inhibitora WDR5

Simulacije MD su izvršene na četiri različita sistema: Sistem A: apo WDR5; Sistem B: WDR5:*cis*-CS-VIP 8; Sistem C: MLL1 peptid-WDR5-RbBP5 peptid; Sistem D: *cis*-CS-VIP 8-WDR5-RbBP5 peptid. Fotoizomerizacija CS-VIP 8 je proučavana korišćenjem a) molekulske dinamičkih simulacija "diabatskog promenljivog potencijala"; b) dobro temperirane metadinamike (MetaD). MD simulacije su izvršene korišćenjem GROMACS 2016.6 softvera²⁹⁵. Početne koordinate su dobijene iz PDB ID: 7AXP, ili iz PDB ID: 3P4F za sisteme koji sadrže RbBP5. Ispitivani sistemi su predstavljeni pomoću AMBERff14SB16 polja sile²⁹⁶, GAFF2 polja sile²⁹⁷ i TIP3P modela vode. Parametri vezivanja azobenzena (uključujući parametre za pobuđeno stanje S1) su preuzeti iz literature²⁹⁸. Parcijalna naelektrisanja azobenzenskog fragmenta su prilagođeni korišćenjem pristupa ograničenog elektrostatičkog potencijala (eng. *Restrained Electrostatic Potential* – RESP) RESP/HF-6-31G*²⁹⁹. Nakon neutralizacije i minimizacije (5000 koraka gradijenta strmog spusta), sistemi su prvo ekvilibrirani u NVT ansamblu (310 K, Nosé-Hoover termostat) a zatim u NPT ansamblu (1.0 bar, Parrinello-Rahman barostat). Pozicione restrikcije su postepeno uklanjane (od 1000 kJ/mol nm⁻²) pre produkcione simulacije. *Particle Mesh Ewald* (PME) je upotrebljen za računanje elektrostatičkih interakcija na dugom rastojanju uz graničnu vrednost od 10 Å. SHAKE algoritam je upotrebljen za ograničavanje veza. Dobijene trajektorije su analizirane korišćenjem ugrađenih Gromacs alata. Analiza skeniranja perturbacionog odgovora (eng. *Perturbation response scanning* – PRS) je izvršena korišćenjem ProDy softvera³⁰⁰. Da bi se identifikovale alosterijske putanje, korišćen je softver gRINN²⁴⁷ i interakcione mreže su dodatno analizirane korišćenjem Cytoscape 3.7.2 program primenom DyNet dodatak i Dn-skora^{248,249}. MD simulacije dijabatskog prebacivanja su izvedene prema adaptiranom protokolu opisanom u literaturi³⁰¹. MetaD simulacije su izvedene korišćenjem PLUMED 2.3.5 softvera u kombinaciji sa GROMACS 2016.6 i CNCDiedarskim uglom kao kolektivnom varijablom²⁵⁷. Faktor pristrasnosti, stopa depozicije, početna visina i širina *Gaussian*-ovog potencijala pristrasnosti su podešeni redom na 25, 0,5 ps⁻¹, 2,5 kJ·mol⁻¹ i 0,35 rad.

3.3. Bioinformatička analiza novih ciljnih mesta za razvoj dualnih inhibitora histon-deacetilaza

Farmako-transkriptomički odnosi između bazalne ekspresije gena i hemijske osetljivosti ćelijskih linija adenokarcinoma pankreasa na odgovarajuće bioaktivno jedinjenje ispitivani su korišćenjem sistematske korelacione analize velikih javno dostupnih skupova podataka: genomičke baze osetljivosti tumorskih ćelija na lekove „GDSC“ (eng. *Genomics of Drug Sensitivity in Cancer* – GDSC) (<https://www.cancerxgene.org/>)³⁰², portala terapijskog odgovora malignih ćelija „CTRP“ (eng. *Cancer Therapeutics Response Portal* – CTRP) (<https://portals.broadinstitute.org/ctrp.v2.1/>)²²² i Enciklopedije tumorskih ćelijskih linija (eng. *Cancer Cell Line Encyclopedia* – CCLE) (<https://depmap.org/portal/download/22Q2>)³⁰³. *Pearson*-ovi koeficijenti korelacije izračunati su između površine ispod krive (AUC vrednosti – mera odgovora ćelijske linije na tretman) i podataka o bazalnoj ekspresiji (izraženoj kao log₂RSEM (RNA-Seq metodom maksimizacije očekivanja)) za svaki transkript u svim testiranim ćelijskim linijama. Matrica *Pearson*-ovih koeficijenata korelacije dalje je normalizovana korišćenjem *Fisher*-ove *z*-transformacije kako bi se prilagodila varijacijama u broju testiranih ćelijskih linija za svaku analiziranu molekulu. Transkripti koji odgovaraju *Tukey*-evim ekstremima (1,5 interkvartilni raspon) za svaku malu molekulu dalje su izdvojeni i sačuvani kao „otisci“ transkriptoma koji ukazuju na rezistenciju ili osetljivost na lekove.

Svaki od transkripta iz otisaka osetljivosti ili rezistencije je dalje evaluiran u pogledu značajnosti pozitivnih ili negativnih koeficijenata korelacije sa svim analiziranim bioaktivnim jedinjenjima. Samo su značajne korelacije (*p*-vrednost < 0,05) bile uzete u obzir i svaka iteracija je zabeležena kao „Broj +/-“ ili „Broj -/+“ preklapanja. Ukupan rezultat, nazvan „*Final_Syn_Score*“, izračunat je sabiranjem ovih brojanja. Lista anotiranih ciljnih mesta za sve detektovane sinergizme (*Final_Syn_Score* > 10 i barem jedan transkript izvan granica *Tukey* ekstrema zajednički za par

bioaktivnih jedinjenja) dalje je analizirana pomoću funkcionalne analize obogaćenja sa g: Profiler servera (<https://biit.cs.ut.ee/gprofiler>)³⁰⁴.

3.3.1. Eksperimentalna validacija sinergizama na MIA Paca-2 i Panc-1 ćelijskim linijama

Ćelije MIA Paca-2 (ATCC CRL-1420) i Panc-1 (ATCC CRL-1469) su kultivisane u DMEM medijumu (Sigma-Aldrich, St. Louis, MO, SAD) koji je dopunjen sa 10% fetalnog goveđeg seruma (FBS) (Sigma-Aldrich, St. Louis, MO, SAD), 100 µg/mL streptomocina i 100 U/mL penicilina (Sigma-Aldrich, St. Louis, MO, SAD). Kulture su održavane kao monosloj u vlažnoj atmosferi sa 95% vazduha i 5% CO₂ na temperaturi od 37°C u inkubatoru sa 5% CO₂.

Sinergistički efekti kombinacija lekova/malih molekula su praćeni MTT testom vijabilnosti uz primenu metode kombinatorijskog indeksa opisanog od strane Choua i Talalay-a³⁰⁵. Citotoksična aktivnost sintetisanih inhibitora HDAC, ROCK inhibitora RKI-1447 (Selleck Co. (Shangaj, Kina)) i Fingolimod HCl (Sigma-Aldrich, St. Louis, SAD) je procenjena na MIA PaCa-2 i Panc-1 ćelijama koristeći MTT test²⁸⁸. MIA PaCa-2 (4×10^3 ćelija/bazenu) i Panc-1 (5×10^3 ćelija/bazenu) su tretirane sa sintetisanim jedinjenjima u pet različitih koncentracija (100, 50, 25, 12.5 i 6.25 µM), pri čemu je svaka koncentracija dodata u pet ponavljanja. Nakon 72 sata, svakoj testiranoj koncentraciji je dodato 20 µL MTT rastvora (3-(4, 5-dimetiltiazol-2-il)-2, 5-difeniltetrazolijum bromid) (Sigma-Aldrich, St. Louis, SAD). Uzorci su inkubirani 4 sata, nakon čega je dodato 100 µL 10% SDS-a i inkubirano na 37 °C. Narednog dana je izmerena apsorbance na 570 nm. Preživljavanje ćelija (%) je izračunato kao odnos apsorbance (pri 570 nm) između tretiranih i kontrolnih ćelija, pomnožen sa 100. IC₅₀ je definisana kao koncentracija agensa koja inhibira preživljavanje ćelija za 50% u poređenju sa kontrolnom grupom.

Kombinatorijski indeks opisan od strane Choua i Talalay-a³⁰⁵ je računat na osnovu podataka iz MTT testa koristeći softver CalcuSyn (verzija 2.0, Biosoft, Kembridž, Velika Britanija). Jedinična vrednosti kombinatorijskog indeksa (CI) ukazuje na aditivni efekat, dok vrednosti CI = 0.3–0.7 ukazuje na sinergizam, CI = 0.1–0.3 ukazuje na snažan sinergizam, a CI < 0.1 ukazuje na veoma jak sinergizam.

4. Rezultati i diskusija

4.1. Ispitivanje konformacione dinamike SIRT2 i razvoj protokola za SBDD dizajn novih SIRT2 inhibitora

Usled nedostatka jasne veze odnosa struktura SIRT2 inhibitora i konformacione heterogenosti unutar kristalografski razrešenih SIRT2:inhibitor kompleksa (Slika 1.14), racionalni dizajn selektivnih SIRT2 inhibitora je započeo detaljnim ispitivanjem konformacione dinamike SIRT2.

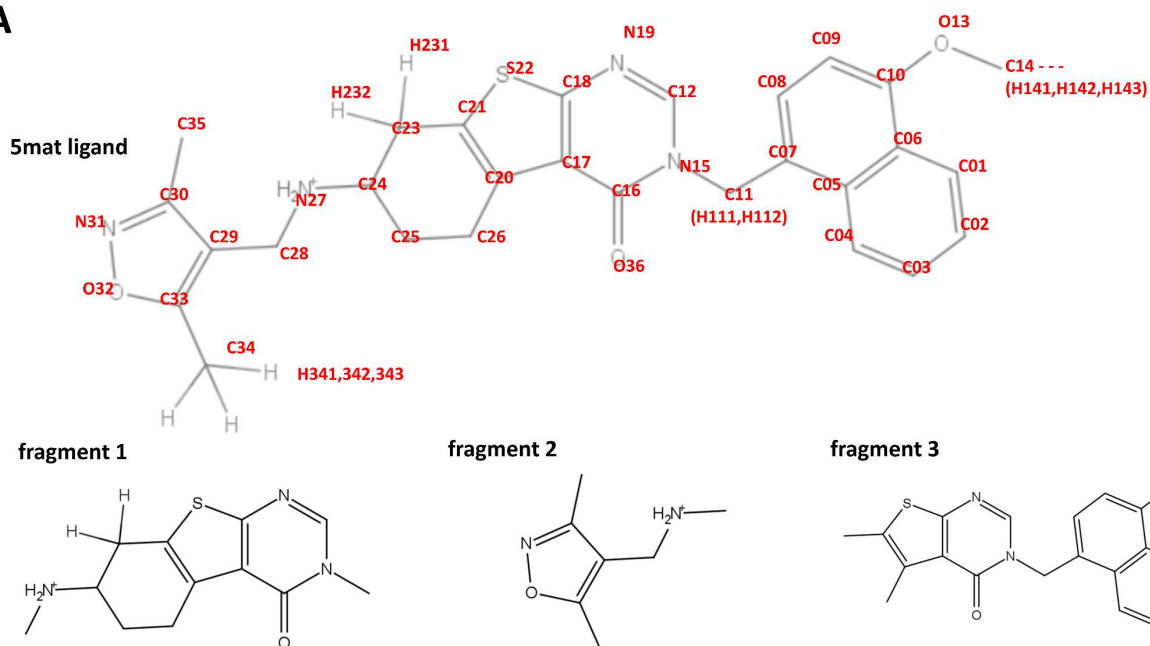
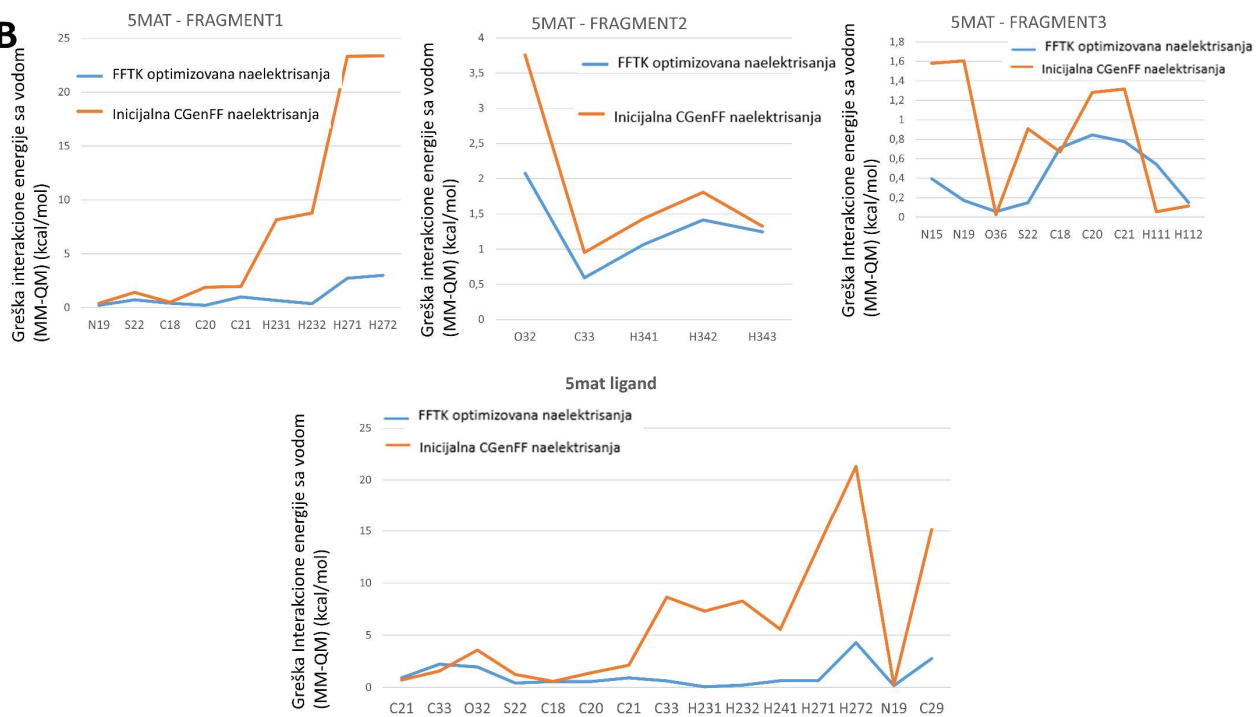
Inicijalno ispitivanje konformacione dinamike SIRT2 u *apo* formi, ili u kompleksu sa inhibitorima, izvedeno je primenom simulacija molekulske dinamike (MD). MD simulacije su izvedene primenom osam različitih startnih konfiguracija SIRT2 (Slika 1.14 i 3.1). Izbor sistema za simulaciju je izveden je na osnovu razlika u konformacionom stanju SIRT2 indukovanom prisustvom inhibitora, uzimajući u obzir i diverzitet hemijskih struktura samih inhibitora. U tom kontekstu su se, pored razlika u načinu vezivanja unutar vezivnog mesta SIRT2, odabrani inhibitori razlikovali i u veličini i volumenu vezivnog mesta koji su zauzimali. Ligand 5yqo bio je najveći među ispitivanim ligandima i jedini koji zauzima celu dužinu supstratnog vezivnog mesta (Slika 1.14). Pri izboru sistema za simulaciju, vođeno je računa da svi odabrani ligandi ostvaruju interakciju sa tzv. džepom selektivnosti SIRT2. Interakcije unutar ovog džepa su prepoznate kao jedno od obeležja selektivnosti SIRT2 inhibitora s obzirom da se ovaj vezivni džep ne otvara u vezivnim mestima inhibitora ostalih sirtuinskih izoformi¹⁸⁰.

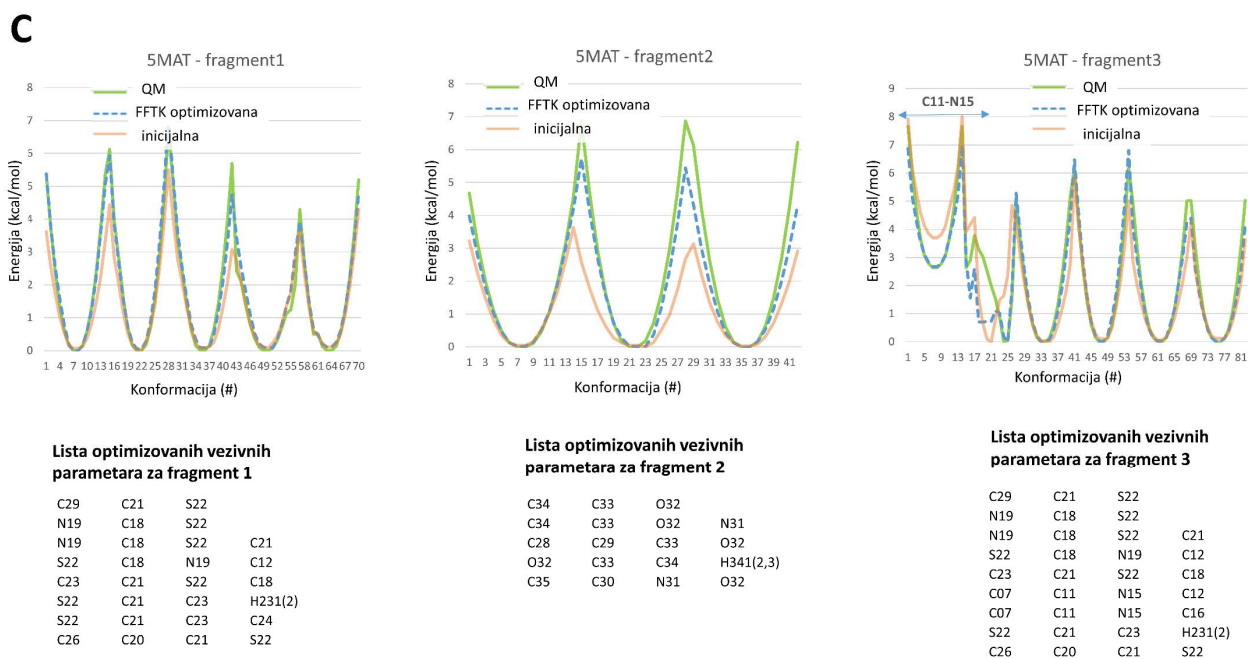
Pre početka simulacija molekulske dinamike i metadinamičkih simulacija, kvalitet dostupnih parametara molekulske mehanike (MM) polja sile za liganade testiran je primenom CGenFF servera. Za većinu ispitivanih liganada, parametrima je dodeljen visok kazneni skor čime je ukazano na problematičan kvalitet, odnosno nepostojanje hemijski dovoljno sličnih liganada u bazama podataka MM parametara. Usled problematičnog kvaliteta postojećih MM parametara liganada, koji mogu biti uzrok greške MD simulacija, izvršena je optimizacija parametara na osnovu QM proračuna primenom ffTK protokola (poglavlje 3.1.1).

4.1.1. Optimizacije molekulske mehanike parametara polja sile SIRT2 liganada

- *Optimizacija parametara sa visokim kaznenim skorom za 5mat ligand*

Imajući u vidu velike zahteve za računarskim resursima QM proračuna za više od 40 atoma na MP2(6-31G*) nivou teorije, 5mat ligand je podeljen na tri fragmenta (Slika 4.1 – fragmenti 1-3). Fragmentacija je izvršena tako da se minimalno naruše hemijske osobine i priroda diedarski uglova fragmenata u odnosu na ceo ligand. Prvo su optimizovana parcijalna naelektrisanja sa većim kaznenim skorovima za fragment 1, a dobijena optimizovana naelektrisanja su dalje koršćena kao početna tačka za optimizaciju parcijalnih naelektrisanja fragmenta 3. Rekonstrukcija celog liganda iz fragmenata izvršena je koristeći standardni pristup opisan u Materijalima i metodama. Međutim, zbog različitih hemijskih osobina fragmenta 1 i 3 (različita naelektrisanja, benzilni ugljenik itd.), rekonstrukcija celog 5mat liganda iz fragmenata izvedena je uz preračunavanje optimizovanih parcijalnih naelektrisanja fragmenata na celom 5mat ligandu. Većina grešaka izračunatih interakcionih energija za optimizovane parametre bila je manja od 1 kcal/mol, što je u skladu sa očekivanom greškom ffTK metode²⁵. Vezivni parametri su optimizovani prvobitno za fragment 1, dok su analogni parametri re-validirani na fragmentu 3. Parametri koji se odnose na torziju oko C11-N15 su posebno prilagođeni zbog odstupanja u distribuciji MM uzorkovanja. Lista optimizovanih parametara je prikazana na Slici 4.1.

A**B**

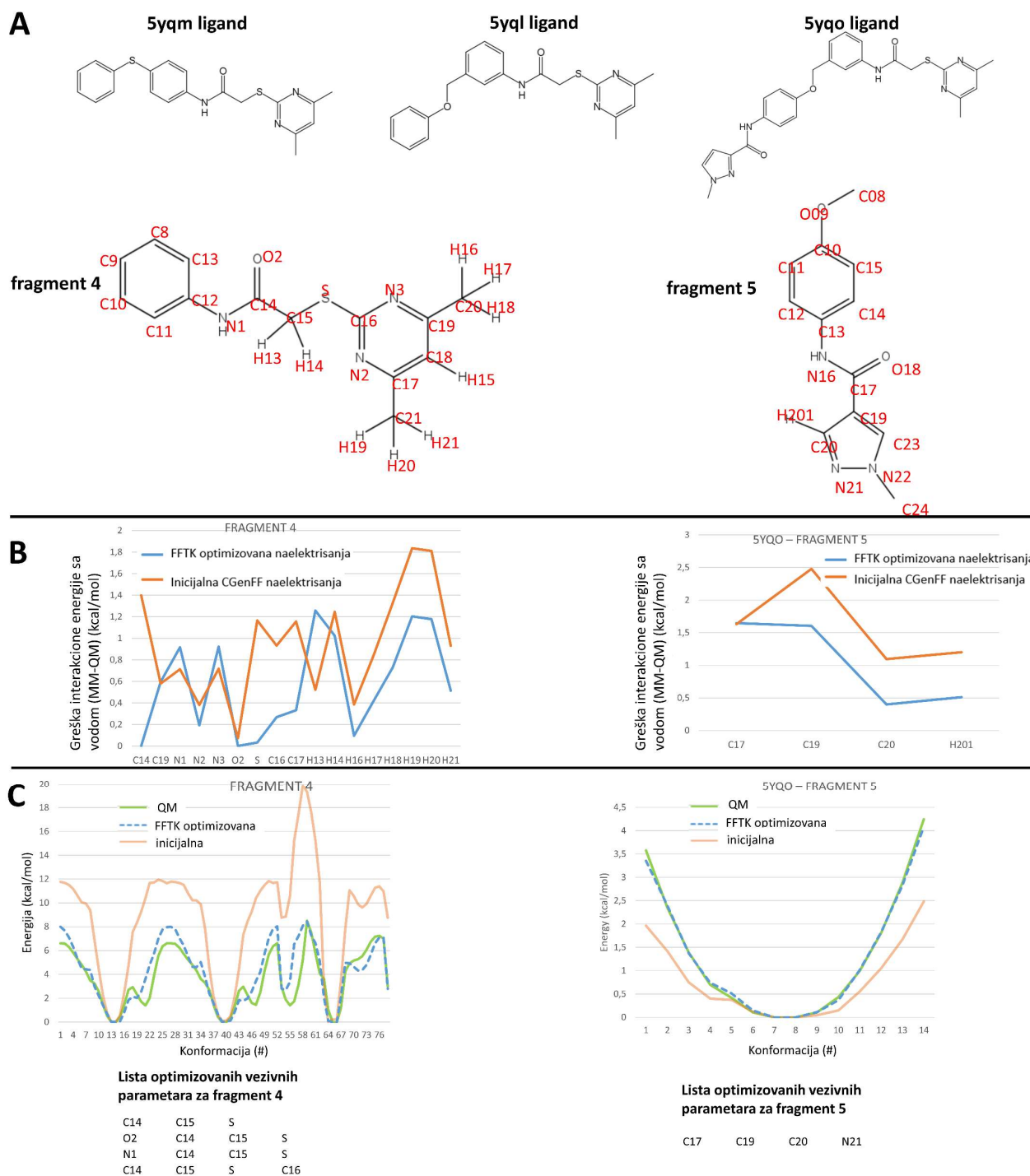


Slika 4.1. Rezultati optimizacije parametara 5mat liganda.

• **Optimizacija parametara sa visokim kaznenim skorom za ligande 5yqo, 5yql i 5yqm**

Za ligande 5yqo, 5yql i 5yqm, najveći kazneni skorovi su detektovani unutar fragmenta 4 (4,6-dimetil-2-pirimidinil)tioacetamida) (Slika 4.2). Većina grešaka izračunate interakcione energije sa TIP3P molekulima vode za optimizovane parametre bila je manja od 1 kcal/mol, što je u skladu sa očekivanim rezultatima fTK optimizacije²⁵. Zbog složene površine potencijalne energije fleksibilne veze C15-S (metastabilno stanje uključujući intramolekulsku H-vezu), izvršena su dva diedarska skeniranja počevši od različitih početnih konfiguracija minimizovanih struktura. Za razliku od polaznih parametara, optimizovani parametri su pokazali visok nivo usklađenosti sa torzionim energijama dobijenim kroz QM proračune (Slika 4.2).

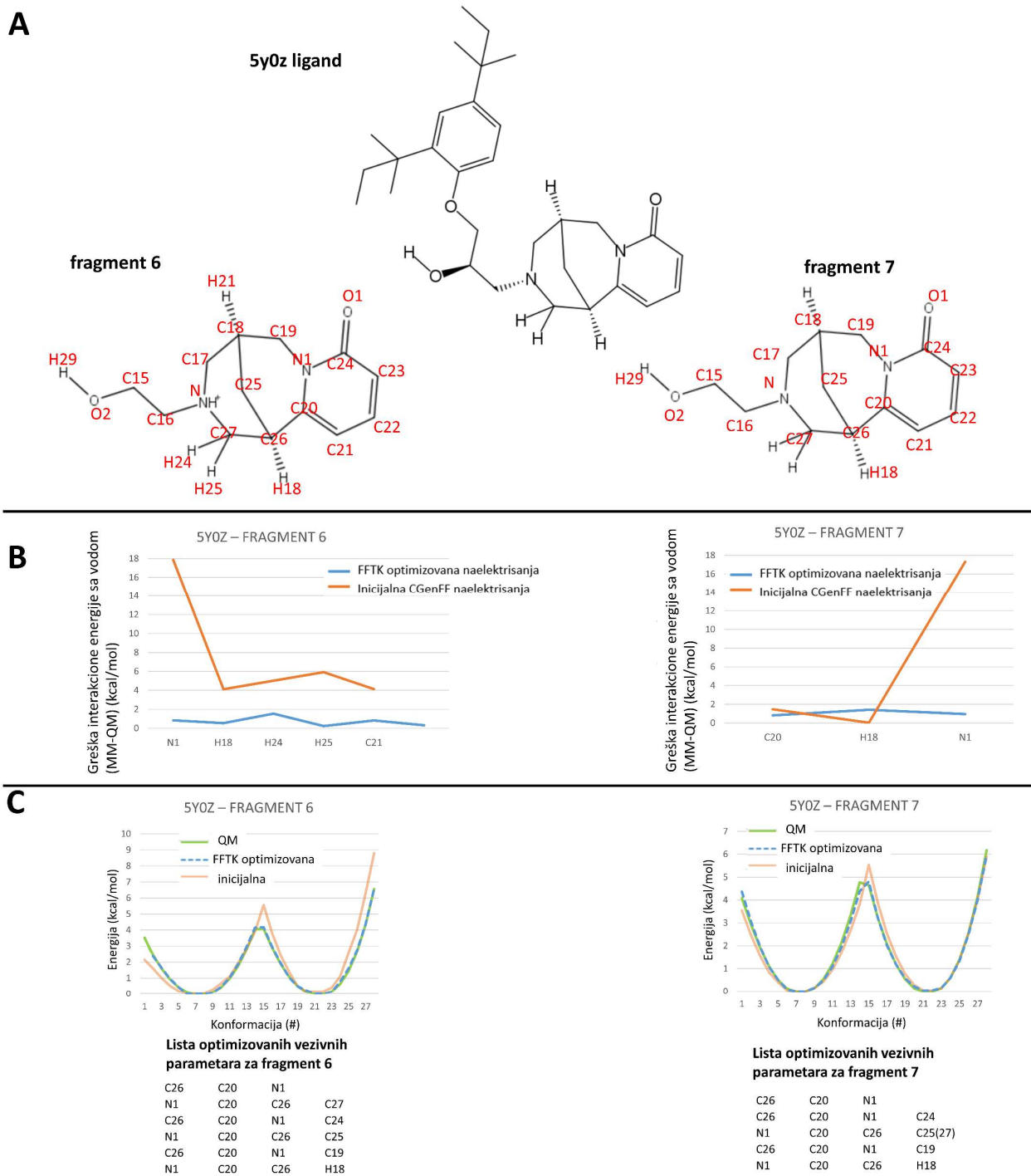
Kod liganda 5yqo, dodatni deo molekula (fragment 5) pokazao je veće kaznene skorove, zbog čega je optimizacija naelektrisanja i diedarskih uglova izvršena prema standardnom protokolu. Celi ligandi (5yqo, 5yqm, 5yql) su rekonstruisani iz fragmenata prema prethodno opisanim protokolima (poglavlje 3.1.1). Lista optimizovanih parametara je prikazana na Slici 4.2.



Slika 4.2. Rezultati optimizacije parametara 5yqo, 5yql i 5yqm liganada.

• **Optimizacija parametara sa visokim kaznenim skorom za ligand 5y0z**

Ligand 5y0z je prema PROPKA 3.11²⁴⁰ predviđen da ima gotovo jednaku raspodelu naelektrisanog i nanaelektrisanog oblika unutar vezivnog mesta. Zbog neizvesnog stanja protonacije unutar vezivnog mesta, ovaj ligand je parametrizovan i simuliran u oba stanja. U oba slučaja je korišćen opisani protokol za optimizaciju. Svi parametri sa visokim kaznenim skorom obuhvaćeni su fragmentima 6 i 7 (Slika 4.3). Optimizacija parcijalnih naelektrisanja rezultirala je značajnim poboljšanjima. Budući da je većina optimizovanih vezivnih parametara vezana za konformaciono ograničenu prstenastu strukturu, optimizovani parametri se nisu mnogo razlikovali od polaznih parametara. Lista optimizovanih parametara je prikazana na Slici 4.3.

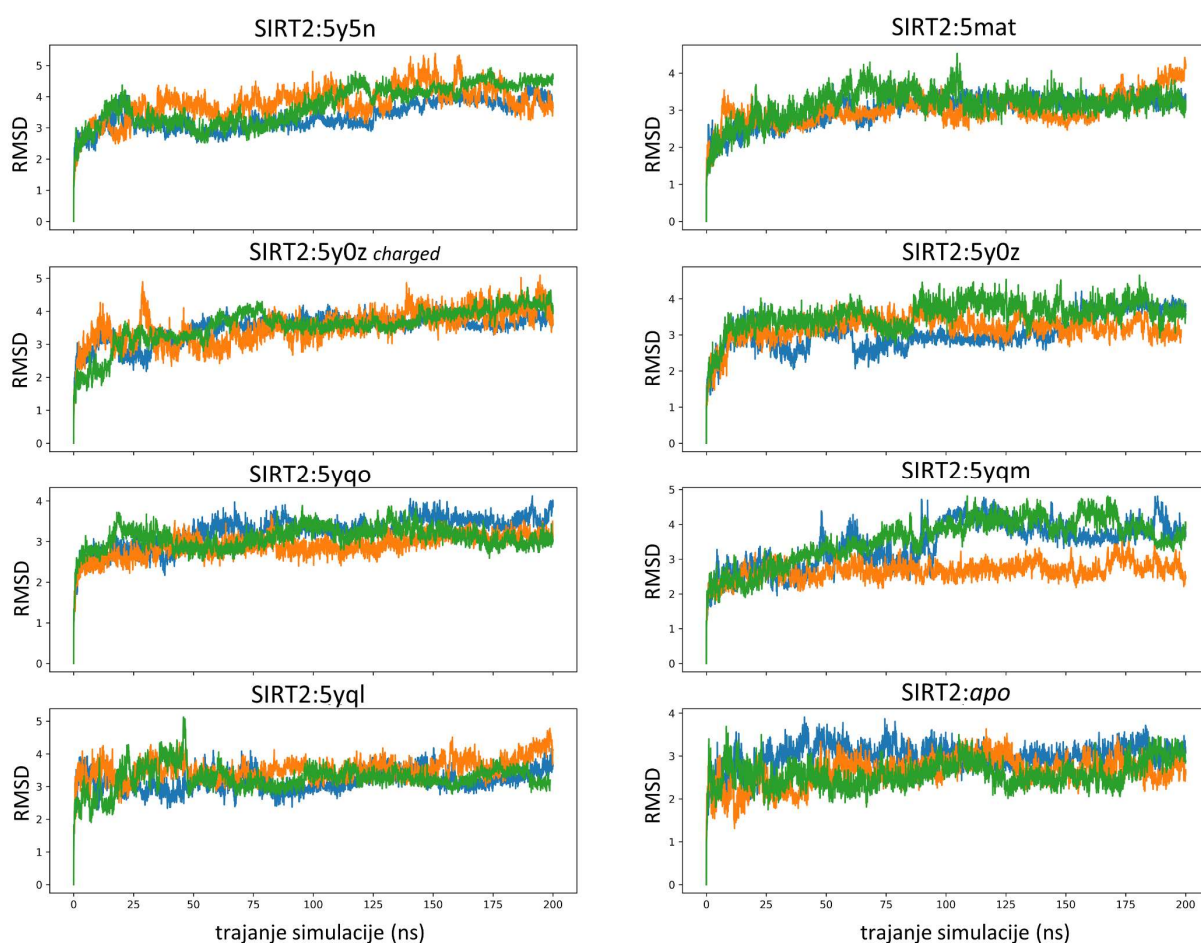


Slika 4.3. Rezultati optimizacije parametara 5y0z liganda u jonizovanom i nejonizovanom obliku.

U zaključku, optimizovani parametri su pokazali veći kvalitet, odnosno veću usklađenost u energijama interakcije sa molekulima vode kao i torzionih potencijala predviđenih primenom QM proračuna. Uzimajući u obzir translatorski potencijal molekulske-mehaničke parametara, značaj optimizovanih parametara može se videti u univerzalnosti funkcionalnih grupa i skeleta koji se pojavljuju u navedenim ligandima što razvijene parametre čini široko primenjivim.

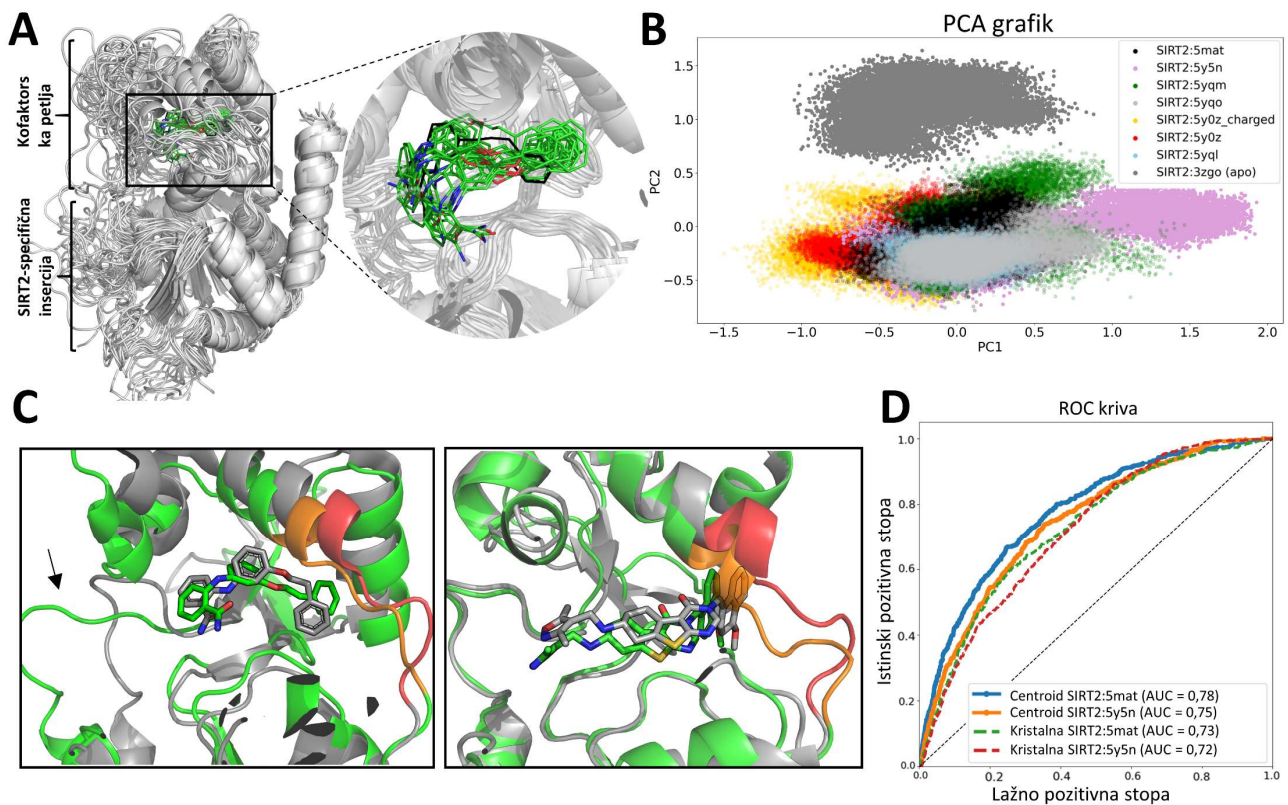
4.1.2. Ispitivanje konformacione dinamike SIRT2 primenom klasične MD simulacije

Svaki od sistema je simuliran kroz 200 ns klasične MD u triplikatima, što čini ukupno 4,8 μ s kumulativnih trajektorija. Svaka od trajektorija je dalje analizirana primenom RMSD analize atoma peptidnog skeleta SIRT2. U okviru svake od simulacija, simulirani ligand je ostao vezan unutar vezivnog mesta dok je RMSD analiza ukazala na konvergirane sisteme bez prisustva bilo kakvih većih konformacionih promena (Slika 4.4). RMSF analiza je ukazala na dva regiona podložna najvećim fluktuacijama: SIRT2 specifična insercija i kofaktorska vezivna petlja (Slika 4.5 i Prilog 7.1). Vizuelni pregled trajektorija ukazao je na različita konformaciona stanja kofaktorske petlje čak i unutar istog simuliranog sistema tokom svakog od tri ponavljanja što ukazuje na fleksibilnu prirodu same petlje i ograničenost uzorkovanja svih konformacionih stepeni slobode tokom 200 ns simulacija (Slika 4.5, Prilozi 7.1 i 7.2). Rezultati RMSF analize bili su u skladu sa prethodno objavljenim rezultatima (Prilog 7.1)¹⁸⁵⁻¹⁸⁷.



Slika 4.4. Rezultati RMSD analize simuliranih sistema. Plava, zelena i narandžasta boja indikuju svaku od replika simuliranog sistema.

Uzimajući u obzir da aminokiseline kofaktorske vezivne petlje PHE-96 i ARG-97 ulaze u sastav građe vezivnog mesta i mogu uspostaviti interakcije sa ligandima, MD simulacije su ukazale na problematiku primene statičnih kristalografski razrešenih struktura u strukturno-zavisnim optimizacijama inhibitora SIRT2. Dodatno, vizuelni pregled MD trajektorija ukazao je na relativno visok broj alternativnih vezivnih stanja (konformacija) samih liganada unutar vezivnog mesta što je bilo naročito izraženo kod manjih liganada poput 5y5n (Slika 4.5A i Prilog 7.2).

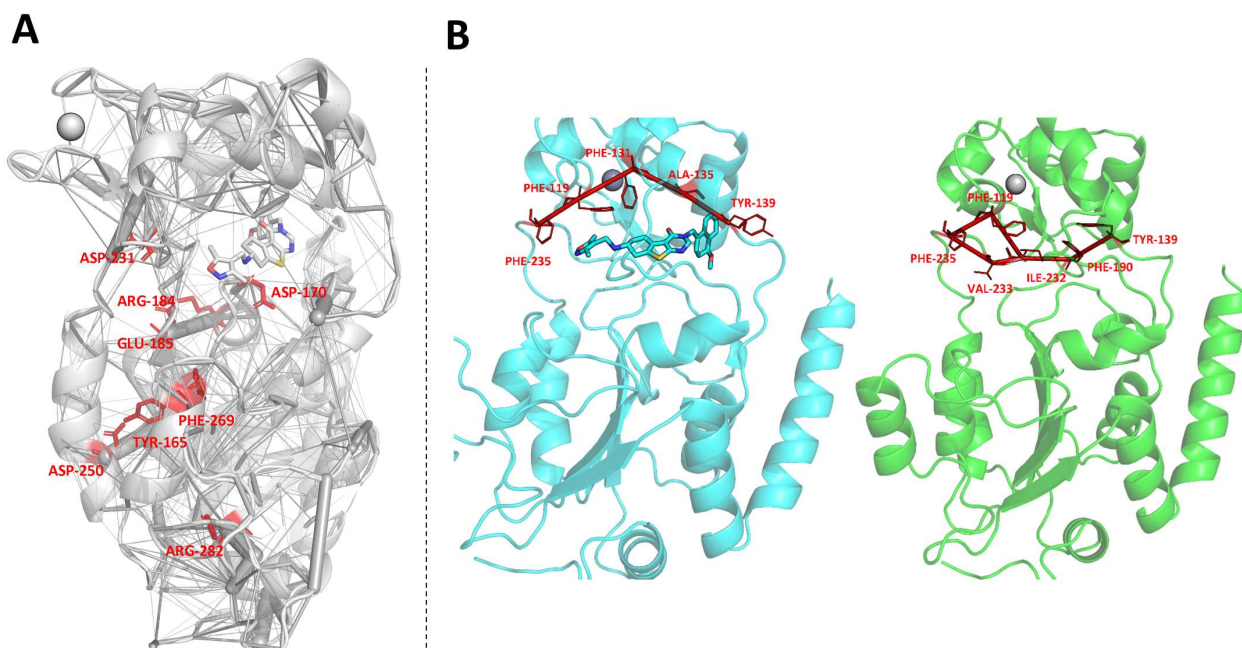


Slika 4.5. Konformaciono ponašanje SIRT2 opisano kroz klasične MD simulacije i validacija opisanih stanja kroz retrospektivni SBVS. A – Ansambl konformacionih stanja SIRT2:5y5n dobijenih kroz klasične MD simulacije (alternativni načini vezivanja liganda prikazani su zelenom bojom, dok je kristalografski ligand prikazan crnim linijama); B – Rezultati PCA analize vezivnog mesta za sve simulirane sisteme; C – Poređenje najbolje rangiranih centroida (zeleno) SIRT2:5y5n (levo) i SIRT2:5mat (desno) sa odgovarajućim kristalografski razrešenim strukturama (siva boja). Petlja 135-143 je obeležena narandžastom (MD centrioda) ili crvenom (kristalografski razrešena) bojom. Crna strelica ukazuje na deformaciju supstratnog vezivnog mesta SIRT2; D – ROC krive dobijene u retrospektivnom SBVS-u na MD centriodima.

U cilju dalje analize veze alternativnih vezivnih stanja samih liganada sa alternativnim konformacionim stanjima vezivnog mesta SIRT2 izvedena je PCA analiza dobijenih trajektorija. PCA analizom utvrđeno je da su uočene razlike u načinu vezivanja liganada u vezi sa alternativnim konformacionim stanjima vezivnog mesta. Očekivano, PCA analiza ukazala je na vrlo jasnu razliku u konformacionom ponašanju *apo* i *holo* stanja (Slika 4.5B).

Analiza dinamičke interakcione mreže rezidua (eng. *Dynamical residue interaction network* – RIN) je dalje upotrebljena radi analize funkcionalno značajnih rezidua SIRT2 enzima i njihove povezanosti sa uočenim konformacionim različitostima između *apo* i *holo* stanja. Kroz RIN analizu, svaka aminokiselina je predstavljena kao čvor, dok su nekovalentne interakcije između njih predstavljene kao veze u mreži. Dinamička mreža interakcija rezidua je generisana za svaki od simuliranih sistema i analizirana primenom centralizovane metrike (eng. *centrality*) i metrike prespajanja (eng. *rewiring*).

Mera uzajamne centralizovanosti (eng. *Betweenness centrality* – BC) je korišćena kao indikator nivoa kontrole koju svaka od rezidua poseduje unutar komunikacione mreže celog SIRT2 enzima. BC analizom izdvojena je rezidua ASP-170 kao jedan od najznačajnijih komunikacionih centara unutar samog enzima. Specifično pozicioniranje ASP-170 (na granici domena *Rossmann*-ovih nabora i Zn^{2+} -vezivnog domena) ukazuje na to da ova aminokiselina predstavlja važno čvorište komunikacije između funkcionalnih domena SIRT2. Rezultati analize uzajamne centralizovanosti su prikazani na Slici 4.6 i u Prilogu 7.3.



Slika 4.6. Rezultati analize dinamičke interakcione mreže rezidua. A – Prikaz rezidua SIRT2 sa najvišim BC vrednostima (crveno) unutar projektovane dinamičke interakcione mreže (sivo); B – Identifikovani komunikacioni putevi (crveno) unutar *holo* (levo) i *apo* (desno) sistema.

Analizom prespajanja uočeno je da se tokom vezivanja liganda značajno menja struktura mreže komunikacije unutar samog SIRT2 pri čemu neke od najznačajnijih promena uključuju rezidue iz petlje 135-143 i reziduu PHE-190 koje ulaze u sastav džepa selektivnosti u okviru vezivnog mesta SIRT2. Aminokiseline sa najvećim skorom primenjene metrike prespajanja (DyNet skor) su prikazane u Tabeli 4.1. Analizom prespajanja, kao najznačajnije razlike u strukturama *apo* i *holo* RIN mreža istakli su se komunikacioni putevi posredovani aminokiselinama iz džepa selektivnosti: aminokiseline iz petlje 135-143 i PHE-190 (označene rezidue Tabeli 4.1). Treba napomenuti da je u *apo* stanju, PHE-190 bila direktno uključena u interakciju sa petljom 135-143, dok je u *holo* stanjima ova interakcija izostala, odnosno bila je premošćena samim ligandom. Nedostatak ove interakcije uticao je na glavne komunikacione puteve između mesta vezivanja supstrata (počevši od aminokiseline PHE-235) i džepa selektivnosti, koji su u slučaju *holo* stanja preusmereni preko aminokiseline PHE-119 (Slika 4.6). Zanimljivo je da su PHE-235, PHE-119 i petlja 135-143 upravo strukturni elementi koji pokazuju najveću konformacionu osetljivost na prisustvo različitih liganda (primećeno u kristalografski razrešenim strukturama) (Slika 1.14). Može se zaključiti da prisustvo različitih hemijskih fragmenata inhibitora u supstratnom vezivnom mestu može imati posledice na konformaciono ponašanje petlje 135-143 putem različitih konformacionih adaptacija PHE-119. Činjenica da je petlja 135-143 direktno uključena u oblik vezivnog džepa, ukazuje na to da bi detaljnije razmatranje konformacione dinamike vezivnog mesta SIRT2 inhibitora dovelo do značajnih benefita u strukturno-zavisnom dizajnu novih jedinjenja sa inhibitornom aktivnošću.

Tabela 4.1. Rezultati analize metrike prespajanja mreže izražene kroz DyNet skor.

5y5n/apo	5mat/apo	5yqo/apo	5yqm/apo	5yql/apo	5y0z_charged/apo	5y0z/apo
143*	301	104	139*	190*	294	301
104	190*	137*	190*	103	301	190*
301	210	103	167	301	137*	137*
139*	97	301	274	167	304	104
110	294	190*	210	139*	303	304
190*	167	167	171	137*	292	303
210	139*	300	143*	143*	190*	139*
299	137*	299	109	106	265	302
140*	104	274	104	104	264	300
107	103	265	95	274	209	292
167	209	139*	303	110	139*	134
109	187	263	301	308	104	309
209	171	187	170	300	300	294
300	131	171	137*	299	288	263
142*	119	106	119	292	263	210
103	96	298	309	210	171	209
171	299	185	304	171	243	171
137*	269	140*	294	142*	229	110
304	229	134	292	140*	210	308
298	142*	110	282	111	131	307
105	134	107	263	96	110	299
101	110	102	233	307	302	298
96	304	100	229	304	298	297
91	292	308	187	298	297	229
307	263	307	149	282	291	185
274	203	303	131	265	282	168
193	201	297	118	263	269	165
134	193	296	107	187	185	143*
269	170	295	105	185	167	142*

* residue vezivnog mesta

4.1.3. Validacija alternativnih konformacionih stanja primenom retrospektivnog virtuelnog skrininga

U cilju potvrde relevantnosti alternativnih konformacionih stanja vezivnog mesta detektovanih kroz klasične MD simulacije, izvedena je studija retrospektivnog strukturno-zavisnog virtuelnog skrininga (SBVS). Alternativna konformaciona stanja su dobijena procedurom klasterovanja trajektorija MD simulacija, pri čemu je svaki centroid identifikovanog klastera okarakterisan kao alternativno konformaciono stanje.

Za potrebe retrospektivnog SBVS, kreirana je baza podataka poznatih inhibitora. Podaci o strukturi i aktivnosti poznatih inhibitora preuzeti su sa ChEMBL baze podataka. Nakon podele jedinjenja na aktivne (513) i inaktivne (114) na osnovu kriterijuma preciziranih u poglavlju 3.1.4 i dodatka molekula virtuelnih mamaca, finalni set podataka za retrospektivni skrining sadržao je 12866 jedinjenja. Vezivna mesta i ligandi su opisani pomoću 4 farmakoforna fingerprinta koristeći GRID-MIF probe (H, N1, O, DRY). Performanse tokom SBVS su praćene kako za kristalografske strukture, tako i za centroide, odnosno alternativna konformaciona stanja SIRT2.

Kao metrika za prećenje performansi upotrebljeni su površina ispod ROC (eng. *Receiver operating characteristic curve*) krive – ROC_AUC i faktor obogaćenja (eng. *Enrichment factor* – EF) ROC-a (EF1%, EF2% i EF5%). Vrednosti EF1%, EF2% i EF5% pokazuju sposobnost SBVS modela da pronade aktivna jedinjenja među najbolje rangiranim rezultatima skrininga u okviru prvih 1, 2, odnosno 5% lažno pozitivnih jedinjenja. Uprkos relativno velikim odstupanjima konformacije vezivnog mesta centroida u odnosu na kristalografski razrešenu strukturu, karakteristike SBVS modela izgrađenih pomoću centroida su bile podjednako dobre, ili čak bolje u odnosu na SBVS modele izvedene iz kristalografske studije. Najveće poboljšanje je zabeleženo u slučaju vrednosti EF faktora gde su VS modeli izvedeni iz SIRT2:5y5n i SIRT2:5mat centroida uspeli da u okviru 2 i 5% najbolje rangiranih lažno pozitivnih rezultata udvostruče broj stvarno pozitivnih jedinjenja u odnosu na kristalografske strukture (Tabela 4.2 i Slika 4.5D). Ovi rezultati potkrepili su inicijalnu hipotezu o neadekvatnosti kristalografski razrešenih struktura SIRT2 u strukturno-zavisnom dizajnu inhibitora.

Tabela 4.2. Rezultati retrospektivnog SBVS.

Grupa	Sistem	ROC_AUC	EF 1%	EF 2%	EF 5%
Kristalografski razrešene strukture SIRT2:inhibitor kompleksa (X-ray)	SIRT2:5mat	0,73	0.03	0.07	0.17
	SIRT2:5y5n	0,72	0.02	0.06	0.15
	SIRT2:5yqo	0,73	0.05	0.1	0.19
	SIRT2:5yql	0,7	0.02	0.05	0.16
	SIRT2:5yqm	0,72	0.03	0.07	0.14
	SIRT2:5y0z	0,69	0.03	0.07	0.16
Centroidi iz MD simulacija	SIRT2:5mat ^a	0,78 (+7%)	0.08 (+67%)	0.14 (+100%)	0.27 (+59%)
	SIRT2:5y5n ^a	0,75 (+4%)	0.04 (+100%)	0.1 (+67%)	0.21 (+40%)
	SIRT2:5yqo ^a	0,74 (+1%)	0.05 (0%)	0.12 (+20%)	0.24 (+26%)
	SIRT2:5yql ^a	0,74 (+6%)	0.02 (0%)	0.06 (+20%)	0.16 (0%)
	SIRT2:5yqm ^a	0,74 (+3%)	0.01 (-67%)	0.04 (-43%)	0.1 (-29%)
	SIRT2:5y0z ^a	0,71 (+3%)	0.02 (-33%)	0.07 (0%)	0.16 (0%)
Strukture kompleksa iz MetaD simulacija	SIRT2:5y5n (holo) ^a	0,75 (+4%)	0.06 (+200%)	0.12 (+100%)	0.24 (+60%)
	SIRT2:5yqo (holo) ^a	0,78 (+7%)	0.06 (20%)	0.12 (+20%)	0.27 (+42%)
	SIRT2:5yqo (apo-like) ^a	0,74 (+1%)	0.05 (0%)	0.09 (-10%)	0.21 (+10%)
Multistrukturalni LDA model	SIRT2:5y5n (MetaD) + SIRT2:5yqo (MetaD) + SIRT2:5yql (X-ray) ^{a,b}	0,83 (+15%/+14%/+19%)	0.06 (+200%/+20%/+200%)	0.17 (+183%/+70%/+240%)	0.33 (+120%/+74%/+106%)

^a U zagradama su istaknuti procenti relativnih poboljšanja specifičnih parametara poređenih sa vrednostima dobijenim za kristalografski razrešene strukture kompleksa (prvi red)

^b Relativna poboljšanja za ansambl model su predstavljena redosledom istaknutim u koloni „Sistem”

Poređenjem struktura najuspešnijih MD centroida sa njihovim kristalografskim parovima, uočeno je da su se centroid i kristalografska struktura uglavnom razlikovali u konformacionom stanju petlje 135-143 koja je uključena u građu džepa selektivnosti (Slika 4.5C). Ova prethodno nepoznata konformaciona stanja SIRT2 ukazala su na visoku konformacionu prilagodljivost džepa

selektivnosti. Poređenjem centroida dobijenih iz simulacije SIRT2:5yqo (koji je zauzimao celokupno vezivno mesto, protežući se od supstratnog vezivnog mesta pa sve do džepa selektivnosti) sa centroidima dobijenim iz SIRT2:5y5n simulacije (manji ligand koji nije bio u stanju da u potpunosti zauzme supstratno vezivno mesto), uočena je značajna distorzija supstratnog vezivnog mesta u slučaju liganda 5y5n (Slika 4.5C). Sa druge strane, u retrospektivnom VS, modeli zasnovani na centroidima iz SIRT2:5y5n simulacije bili su jedni od najboljih (Tabela 4.2) što je navelo na zaključak da supstratno vezivno mesto nije ključno mesto, odnosno interakcije na nivou ovog mesta nisu ključne za objašnjenje aktivnost poznatih inhibitora.

Uzimajući u obzir problem uzorkovanja klasične MD, kao i visoku fleksibilnost petlji na nivou vezivnog mesta SIRT2 detektovanu kroz MD simulacije, može se pretpostaviti da postoje i dodatna konformaciona stanja relevantna za dizajn inhibitora koja nisu bila vidljiva tokom klasične MD.

4.1.4. Razvoj protokola za poboljšano uzorkovanje konformacionih stanja SIRT2 primenom metadinamičkih simulacija

Značajna poboljšanja u performansama retrospektivnog VS nastala usled uračunavanja alternativnih konformacionih stanja SIRT2 identifikovanih kroz simulacije MD inspirisala su dalji rad na razvoju strategije za poboljšano uzorkovanje dinamike vezivnog mesta SIRT2.

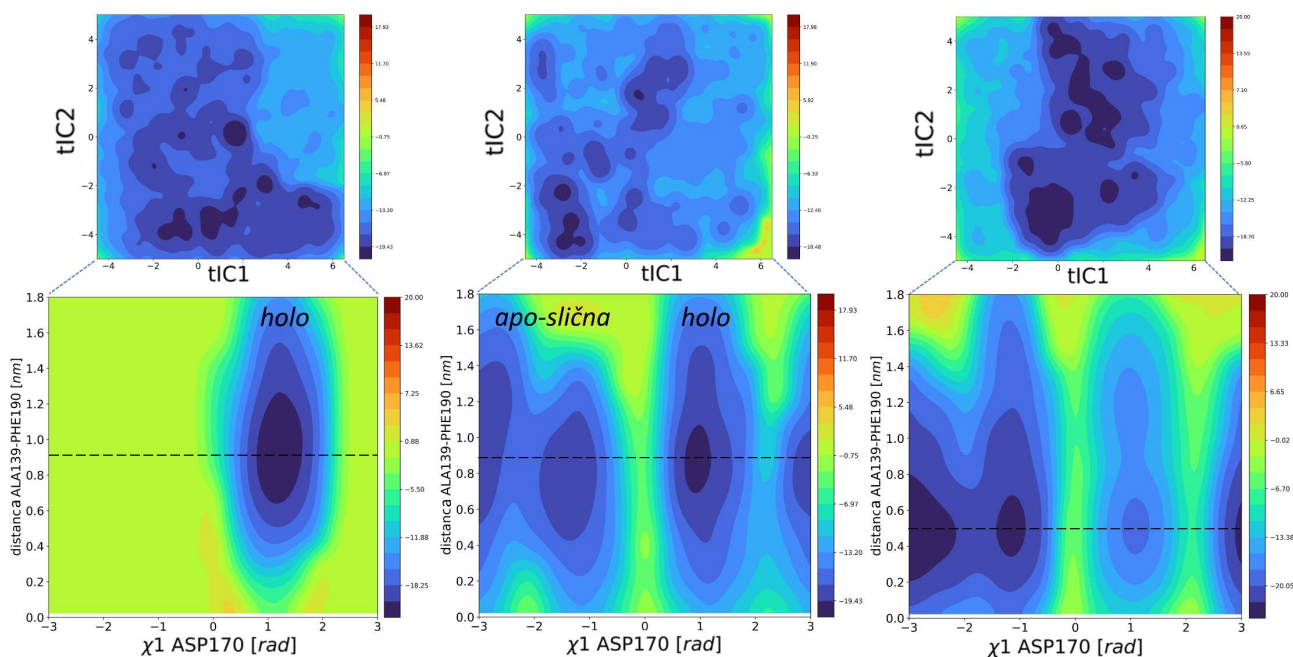
Kao što je već diskutovano u poglavlju 1.2.4, pravilan izbor kolektivnih varijabli (CV) predstavlja jednu od kritičnih tačaka u primeni metadinamike. Pravilno odabrane CV treba da budu sposobne da navode sistem preko svih metastabilnih stanja i tako istraže celokupan fazni prostor omogućavajući ubranu konvergenciju samih simulacija. Tokom ove faze istraživanja, fokus je bio na razvijanju CV pogodnih za efikasnu pretragu različitih konformacionih stanja vezivnog mesta SIRT2 relevantnih za budući razvoj novih inhibitora. Nedavno predložena metodologija tzv. tICA-Metadinamika²⁵⁴, razvijena u cilju ubranog uzorkovanja sporih modova dinamike celokupnih proteina, pokazala je određenu primenu u istraživanju velikih konformacionih promena proteina, ali nikada nije istražen njen potencijal u praćenju dinamike samog vezivnog mesta niti potencijal implementacije u standardne protokole racionalnog dizajna. U okviru istraživanja ove doktorske disertacije, predložena je modifikacija navedenog protokola koja je omogućila poboljšano uzorkovanje dinamike vezivnog mesta i integraciju konformacija relevantnih za racionalni dizajn liganada u SBVS. Fokusiranjem CV na nivo vezivnog mesta, teorijski bi se postiglo efikasnije istraživanje alternativnih konformacionih stanja relevantnih za dizajn inhibitora. Treba skrenuti pažnju da je protokol primenjen u cilju otkrića novih grupa SIRT2 inhibitora sam po sebi univerzalan i može se primeniti na bilo koji ciljani molekul od interesa.

Analiza nezavisnih komponenti kašnjenja (eng. *Time-lagged independent component analysis* – tICA) je tehnika za procesuiranje signala kojom se redukuje dimenzionalnost podataka odgovarajućeg signala dok se minimizuje gubitak kinetičkih informacija (pogledati poglavlje 3.1.3). Sa aspekta molekuskog modelovanja, tICA je najveću primenu pronašla u Markovljevom modelovanju sporih procesa (konformacionih promena) biomakromolekula^{306,307}. tICA može da identifikuje linearnu kombinaciju ulaznih parametara koji maksimizuju vremensku-autokorelacionu funkciju uz fiksno vreme kašnjenja³⁰⁸. Dokazano je da su tICA svojstveni vektori (eng. *eigenvectors*) – tICs sa najvećim svojstvenim vrednostima (eng. *eigenvalues*) zapravo povezani sa najsporijim kolektivnim stepenima slobode, odnosno sporim konformacionim promenama, što opravdava njihovu primenu kao CV u metadinamici²⁵⁴. Treba istaći da je PCA metoda srodna tICA metodi, ali za razliku od nje, PCA traži ortogonalne projekcione vektore (glavne komponente, PC) koji objašnjavaju najveću varijansu unutar trajektorije pritom ne računavajući kinetičku informaciju^{309,310}. Unutar proteina pokreti koji imaju visoku varijansu nisu uvek oni koji su povezani sa sporim pokretima sistema – konformacionim promenama. Primer navedenog slučaja bi bile fluktuacije neuređenih petlji unutar proteina koje se dešavaju na vrlo malim vremenskim skalama. Kofaktorska vezivna petlja SIRT2, kako je pokazano u klasičnim MD simulacijama (poglavlje 4.1.2), predstavlja primer ovakvih pokreta zbog čega primena glavnih komponenti kao

kolektivnih varijabli može da previdi spore pokrete unutar SIRT2 i da favorizuje pokrete kofaktorske petlje. Zbog svega navedenog, primena tICs kao kolektivnih varijabli teorijski predstavlja bolji izbor za studije konformacionih promena unutar SIRT2.

U okviru ovih istraživanja, tICA je izvedena na celom setu trajektorija dobijenih iz klasične MD koristeći *phi* i *psi* diedarske uglove rezidua unutar definisanog vezivnog mesta. Prve dve izdračunate nezavisne komponente (tIC1 i tIC2) opisale su najsporije stepene slobode unutar podataka dobijenih iz MD simulacija. tIC1 i tIC2 su dalje upotrebljene kao reakcione koordinate kako bi ubrzale prelazak sistem preko više različitih konformacionih stanja opisanog najsporijim stepenima slobode. Metadinamička simulacija je izvedena na tri sistema – SIRT2:apo, SIRT2:5y5n i SIRT2:5yqo. Simulacija je trajala do momenta dostizanja kvalitativnog kriterijuma konvergencije (Prilog 7.4). Selekcija *holo* sistema za simulacije je izvedena na osnovu razlika u veličini i hemijskoj strukturi inhibitora kao i načinu interakcije sa vezivnim mestom. Dok je 5y5n mali i fleksibilan ligand koji zauzima manji volumen vezivnog mesta, 5yqo je bio najveći među izučavanim ligandima što mu omogućava da interaguje sa svim delovima vezivnog mesta.

U cilju lakšeg izdvajanja biološki relevantnih konformacionih stanja vezivnog mesta iz rezultata metadinamičkih simulacija, površine slobodnih energija su rekonstruisane pomoću potencijala pristrasnosti deponovanog tokom simulacija metadinamike i projektovane na dve nezavisne komponente – tIC1 i tIC2. Treba istaći poslednja istraživanja ukazuju na oprez prilikom primene tICs u analizi površina slobodne energije, jer tICs često dovode do konstrukcije površina slobodne energije koje su teške za fizičku interpretaciju. Neretko se dešava da se u istom energetskom minimumu nalaze potpuno oprečna konformaciona stanja^{253,308}. U svakom slučaju, projekcijom dobijenih potencijala na nezavisne komponente jasno je uočeno da su profili površina slobodnih energija veoma razlikuju kod oba simulirana *holo* sistema što može ukazati na drugačije ponašanje SIRT2 u zavisnosti od prirode liganda (Slika 7.4).



Slika 7.4. Rezultati metadinamičke simulacije. Gornji grafici predstavljaju projekciju površina slobodne energije na dve nezavisne komponente, dok donji grafici predstavljaju projekciju površina slobodne energije na odabrane kolektivne varijable.

U daljim nastojanjima da se iz metadinamičkih simulacija ekstrahuju konformaciona stanja vezivnog mesta SIRT2 relevantna za dizajn novih inhibitora, pristupilo se proceduri ponderisanja metadinamičkih potencijala na novom setu fizički relevantnih kolektivnih varijabli. Kroz preliminarna promatranja metadinamičkih i MD trajektorija, uočeno je drugačije ponašanje ASP-

170 u *apo* i *holo* strukturama. Uzimajući u obzir diskutovani značaj ASP-170 u transferu signala kroz mrežu interakcija rezidua unutar SIRT2, diedarski uglovi ASP-170 su odabrani za dalja testiranja u analizi fizički relevantnih kolektivnih varijabli. Preliminarnim ispitivanjima uočeno je da *chil* diedarski ugao rezidua ASP-170 najbolje oslikava razliku između *apo* i *holo* stanja. U tom kontekstu ova rezidua može predstavljati senzor prisustva liganda u vezivnom mestu zbog čega je odabran kao prva kolektivna varijabla za ponderisanje (CV1). Naime, ASP-170 je u *apo* strukturama konformaciono manje ograničena, dok u *holo* strukturama *chil* diedarski ugao ove rezidua najčešće ima vrednost od 1 rad (Slika 7.4).

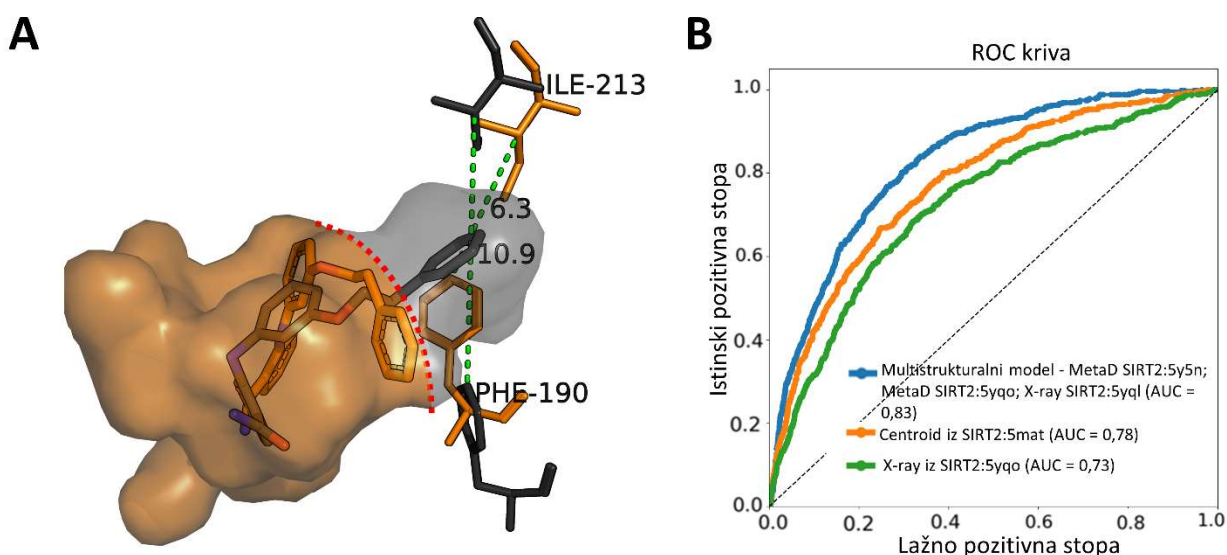
Dalje, uzimajući u obzir potvrđeni značaj interakcija inhibitora sa džepom selektivnosti radi ostvarivanja selektivne inhibicije SIRT2, distanca između rezidua u okviru džepa selektivnosti je odabrana kao druga kolektivna varijabla (CV2) za ponderisanje metadinamičkih potencijala. Rezidua PHE-190 i ALA-135 (iz petlje 135-143 diskutovane u rezultatima klasične MD) predstavljaju ulazna vrata džepa selektivnosti i njihova međusobna distanca direktno oslikava konformaciono stanje pogodno vezivanju selektivnog inhibitora. Ova distanca može biti korisna u praćenju otvaranja džepa selektivnost u *apo* strukturama, kao i u indirektnom praćenju volumena, ili promera džepa selektivnosti.

U poredjenju sa tIC koordinatama tokom simulacija metadinamike, rezultati ponderisanja prikazali su mnogo jasniju sliku dinamičkih promena na nivou vezivnog mesta tokom simulacija metadinamike. Uočeno je da je SIRT2:5y5n metadinamička simulacija dominantno ostala zarobljena na CV1/CV2 površini u poziciji koja oslikava stanja nazvano „potpuno *holo* stanje“ (Slika 7.4). Potpuno *holo* stanje je stanje koje se karakteriše potpuno otvorenim džepom selektivnosti (CV2 vrednosti između 0,8 i 1,0 nm) kao i ograničenom pokretljivošću ASP-170 (oslikano kroz CV1 vrednost od 1 rad). SIRT2:5yqo metadinamička simulacija takođe je dominantno ostala zarobljena u potpunom *holo* stanju, ali su se u slučaju ove simulacije pojavljivala i manje verovatna stanja koja su nazvana „*apo* – slična“ stanja. Ova stanja koja su slična *apo* stanju karakteristična su po većoj pokretljivosti ASP-170 (CV1 zauzima i ostale vrednosti osim 1 rad) karakterističnim za *apo* simulacije, ali i punoj otvorenosti džepa selektivnosti (CV2 vrednosti između 0,8 i 1,0 nm) koji se znatno ređe viđa kod *apo* simulacija. U skladu sa tim, postavljena je hipoteza da *apo* – slična stanja predstavljaju prelazna stanja između *apo* i *holo* struktura. SIRT2:*apo* simulacija je dominantno bila zarobljena u *apo* stanjima koja su okarakterisana zatvorenim džepom selektivnosti, kao i većom fleksibilnošću ASP-170. Interesantno je bilo uočiti da je SIRT2:*apo* metadinamička simulacija bila u stanju da poseti i *holo* stanja (CV1 = 1 rad; CV2 = 0,8–1,0 nm), ali su takvi događaji tokom simulacije okarakterisani kao manje verovatni. Ipak, ovi rezultati ukazuju na to da je proces interakcije SIRT2:inhibitor sa mehanističke tačke gledišta delom vođen konformacionom selekcijom. Uzimajući u obzir sve navedene rezultate, dodatno može biti hipotezirano da je konformaciona rigidifikacija ASP-170 uzrokovana ulaskom liganda u vezivno mesto i zauzimanjem *apo* – sličnih stanja, jedan od pokretača daljeg mehanizma indukovanog fitovanja kojim se SIRT2 finim pokretima specifično prilagođava malom molekulu od interesa i prelazi u puno *holo* stanje.

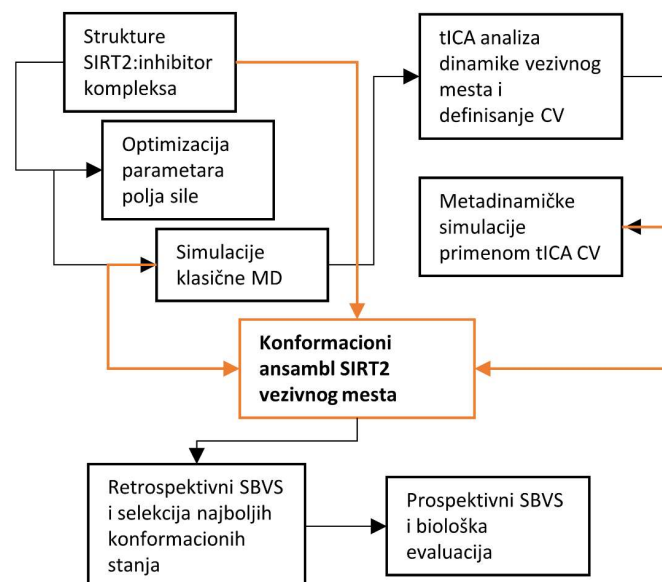
Konformaciona stanja vezivnog mesta SIRT2 najrelevantnija za dizajn novih inhibitora dalje su odabrana uzimajući u obzir *holo* stanja sa projekcije površine slobodne energije na CV1/CV2. U energetskom minimumu koji odgovara *holo* stanjima uočen je čitav niz konformacionih fluktuacija rezidua unutar samog vezivnog mesta. Jedna od najinteresantnijih promena detektovana analizom konformacionog ansambla *holo* stanja jeste otvaranje skrivenog džepa na nivou vezivnog mesta SIRT2 (Slika 4.8). Ovaj skriveni džep se nalazio u nastavku džepa selektivnosti i njegovo otvaranje nije bilo uočeno tokom klasičnih MD simulacija. Otvaranje skrivenog džepa je okarakterisano povećanjem distance između rezidua ILE-213 i PHE-190. Razmatrajući dobro dokumentovan značaj džepa selektivnosti u postizanju selektivnosti inhibicije SIRT2, novootkriveni skriveni džep u nastavku džepa selektivnosti može predstavljati novu osnovu za racionalni dizajn potpuno novih hemijskih grupa SIRT2 inhibitora poboljšane selektivnosti.

4.1.5. Implementacija alternativnih konformacionih stanja SIRT2 sa novootkrivenim skrivenim džepom u protokol SBVS

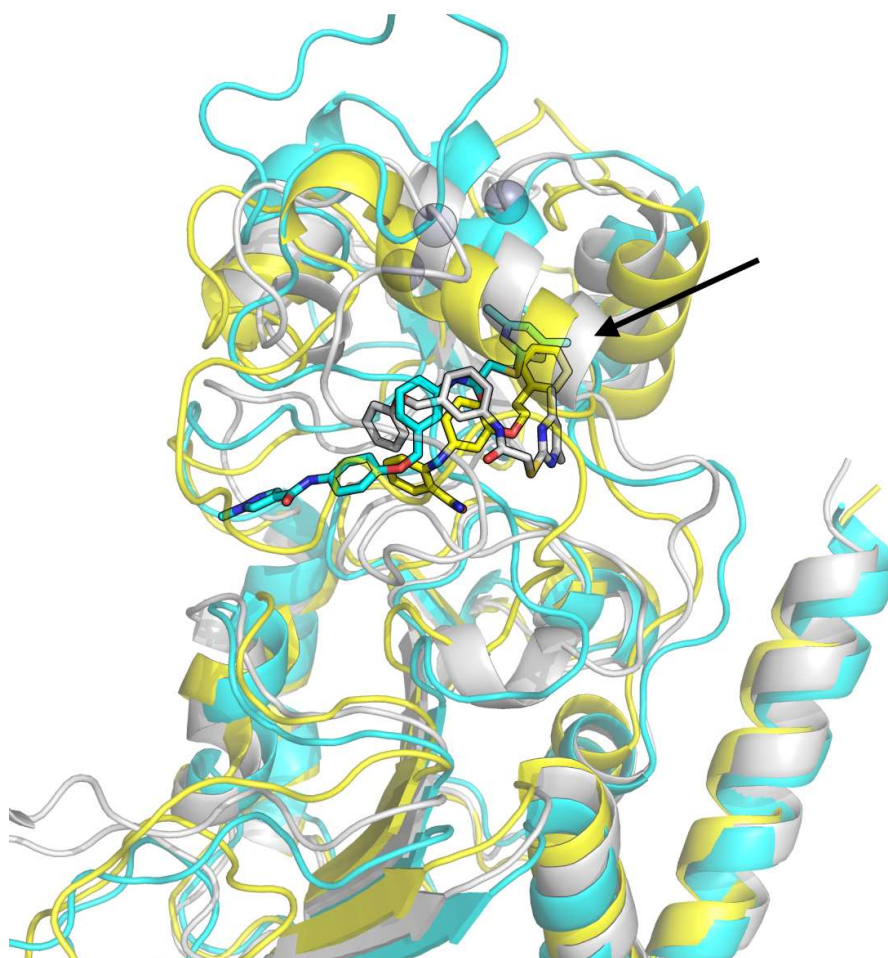
U cilju dalje potvrde relevantnosti novootkrivenih konformacionih stanja kao i skrivenog džepa, prethodno upotrebljen protokol retrospektivnog SBVS je ponovljen (Slika 4.9). Iako su se konformaciona stanja iz metadinamičkih simulacija veoma razlikovala od kristalografskih struktura i struktura ekstrahovanih iz MD simulacija (Slika 4.10), neki od VS modela baziranih na konformacionim stanjima SIRT2 iz metadinamike pokazali su izrazito poboljšanje u performansama SBVS (Tabela 4.2). Konformaciono stanje dobijeno iz SIRT2:5y5n metadinamičke simulacije pokazalo je najveće relativno poboljšanje u EFs metrici od čak 200% u odnosu na kristalografsku strukturu sa istim ligandom. Kao drugi najbolji SBVS model, izdvojen je model zasnovan na SIRT2:5yqo metadinamičkoj simulaciji. Značajno je spomenuti da su oba SBVS modela najviših performansi bila zasnovana na stanjima SIRT2 sa otvorenim skrivenim džepom što doprinosi validaciji relevantnosti novootkrivenog džepa. Interesantno je spomenuti i performanse *apo*-sličnih struktura ekstrahovanih iz metadinamičkih simulacija. Iako su performanse VS modela zasnovanog na ovom konformacionom stanju bile lošije, statistički parametri su ukazali da ovi modeli značajno doprinose objašnjavanju aktivnosti unutar seta podataka što ukazuje na njihovu relevantnost pri interakciji sa inhibitorima.



Slika 4.8. Prikaz novootkrivenog skrivenog džepa (A) i rezultata retrospektivnog SBVS (B). A – Skriveni džep je označen sivom bojom, dok su konformaciono stanje liganda i položaj ključnih aminokiselinskih rezidua označeni crnom. Kristalografska struktura je označena narandžastom bojom; B – ROC krive su prikazane za najbolje SBVS modele unutar svake od grupa prezentovanih u Tabeli 4.2 (kristalografska struktura/MD simulacije/MetaD simulacije).



Slika 4.9. Finalni protokol studije za SBVS zasnovan na alternativnim konformacionim stanjima.



Slika 4.10. Prikaz konformacionih stanja SIRT2 selektivanih za finalni multistrukturalni SBVS model. Tirkiznom i žutom bojom su označena konformaciona stanja selektivana iz MetaD simulacija, dok je sivom bojom označena kristalografska struktura. Crnom strelicom je označen položaj novootkrivenog džepa vidljivog samo u MetaD simulacijama.

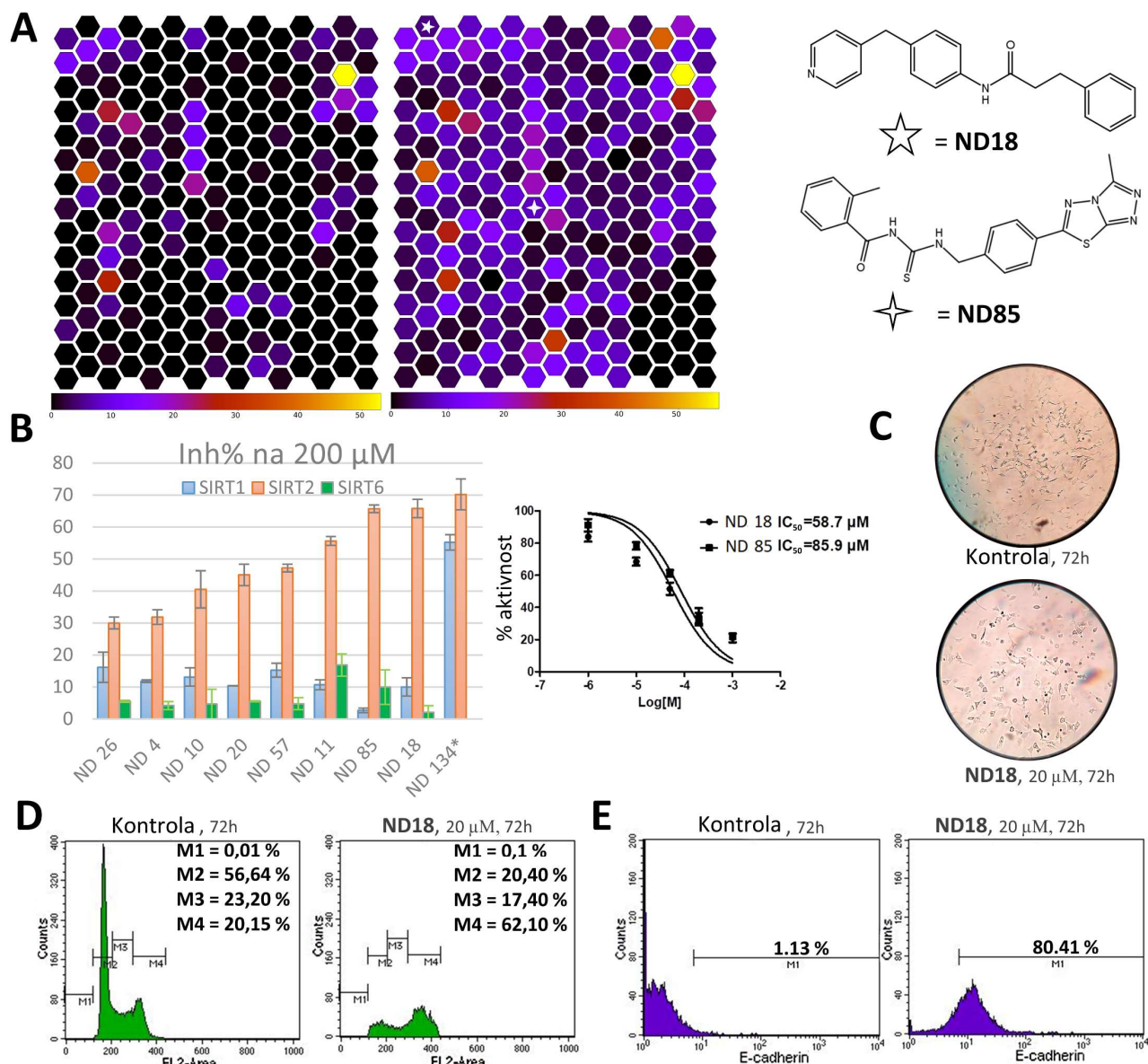
Uz veliku raznolikost konformacionih stanja otkrivenih kroz simulacije, kompleksno konformaciono ponašanje SIRT2 pri interakciji sa inhibitorima ukazuje na scenario u kom je razumno pretpostaviti da bi SBVS sa ansamblom konformacionih stanja SIRT2 pokazao najbolje performanse u studiji retrospektivnog VS. Ansambl konformacionih stanja bi u ovom slučaju mogao da aproksimira doprinos svakog od pojedinačno izdvojenih konformacionih stanja ukupnoj energiji vezivanja inhibitora. Primenom linearne diskriminantne analize (LDA) u okviru FLAP softvera, doprinos alternativnih konformacionih stanja je aproksimiran kroz lineranu kombinaciju konformacija i skorova koji najbolje opisuju varijabilnost unutar trening seta. LDA alat implementiran u FLAP je u automatskom maniru iz skupa konformacionih stanja SIRT2 (uključujući kristalografske konformacije i konformacije iz simulacija) odabrao kombinaciju od 3 konformaciona stanja SIRT2 i 2 skora koja daju najbolje rešenje klasifikacionog problema. Bitno je naznačiti da je LDA alat zapravo selektovao dve strukture iz metadinamike sa otvorenim skrivenim džepom, jednu kristalografsku strukturu i nijednu MD strukturu (Slika 4.10). Multistrukturalni skrining model pokazao je bolje performanse od bilo kog prethodno izgrađenog VS modela (Tabela 4.2 i Slika 4.8). Sveukupno, ovaj rezultat dodatno potkrepljuje značaj novootkrivenog džepa u dizajnu novih inhibitora. Usled najboljih performansi multistrukturalnog modela, ovaj model je dalje selektovan za studiju eksperimentalne validacije.

4.1.6. Eksperimentalna validacija multistrukturalnog SBVS modela kroz ispitivanje neistraženih delova hemijskog prostora

Mogućnost proširenja poznatog hemijskog prostora SIRT2 inhibitora kroz implementaciju novootkrivenih konformacionih stanja SIRT2 vezivnog mesta (predstavljenih kroz multistrukturalni VS model) dalje je ispitana prospektivnim skriningom i *in vitro* validacijom. Primenom multistrukturalnog SBVS izveden je virtuelni skrining preko 200000 komercijalno dostupnih jedinjenja dostupnih u SPECS bazi. Nakon prospektivnog skrininga, strukture približno 1000 najbolje rangiranih jedinjenja (0.5% svih jedinjenja) su izdvojene i zajedno sa poznatim inhibitorima ChEMBL baze enkodirane u obliku MACCS fingerprinta. Dobijeni fingerprintovi su upotrebljeni za kreiranje modela hemijskog prostora SIRT2 inhibitorima primenom Kohonenovih samoorganizujućih mapa. Rezultati analize hemijskog prostora ukazali su na to da je multistrukturalni VS model bio u stanju da istraži prethodno neistražene delove hemijskog prostora SIRT2 inhibitora (Slika 4.11A). U cilju eksperimentalne validacije navedene tvrdnje, devet jedinjenja iz delova hemijskog prostora koji nisu bili adekvatno reprezentovani ChEMBL bazom su nasumično selektovani i testirani *in vitro* na enzimsku aktivnost (Prilog 7.5). Primenom *in vitro* eseja, sva testirana jedinjenja pokazala su značajnu inhibiciju SIRT2 i visoku selektivnost ka SIRT2 (u odnosu na testirane SIRT1 i SIRT6 izoforme) (Slika 4.11B).

Rezultati potvrđuju sposobnost primenjenog SBVS modela da uspešno prepozna jedinjenja sa SIRT2 inhibitornim potencijalom. Strukture devet testiranih jedinjenja sa SPECS ID-om prikazane su na Slici 4.11 i u Prilogu 7.5. Jedinjenje ND134 (SPECS ID: AP-970/43202520) ometalo je fluometrijski esej, što je ograničilo njegovu dalju karakterizaciju *in vitro*. Imajući u vidu da je novootkriveni džep pronađen u produžetku džepa selektivnosti SIRT2, rezultati *in vitro* evaluacije selektivnosti (Slika 4.11B) dodatno potvrđuju važnost uključivanja novootkrivenih SIRT2 konformacionih stanja u SBVS studije. Kako bi se dodatno potvrdila upotrebljivost alternativnih konformacionih stanja SIRT2 u otkrivanju novih hemijskih struktura inhibitora, rangiranje devet selektovanih jedinjenja dobijenih pomoću finalnog multistrukturalnog SBVS modela upoređeno je sa rangiranjem istih jedinjenja nakon ponavljanja prospektivnog skrininga upotrebom SBVS modela baziranih samo na kristalografskim SIRT2 konformacijama. Rezultati su pokazali da je svih devet jedinjenja bilo rangirano značajno ispod unapred definisanog praga od 0,5% (oko 1000 najbolje rangiranih jedinjenja) u slučajevima kada su korišćene kristalografske konformacije (Prilog 7.6). Na kraju, za dva najperspektivnija jedinjenja (ND18 (SPECS ID: AO-080/40098868) i ND85 (SPECS ID: AP-970/43202520)), utvrđene su vrednosti IC₅₀ (SIRT2) praćenjem odnosa doze i aktivnost (Slika 4.11B). Oba jedinjenja pokazala su se kao potentni i selektivni inhibitori SIRT2, sa

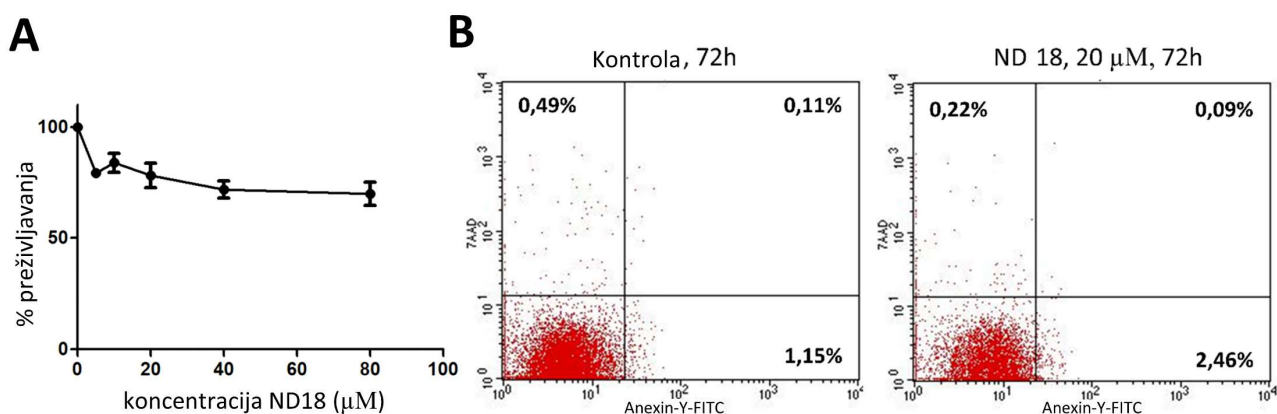
potentnošću koja se može porediti sa dobro poznatim inhibitorom kambinolom (IC_{50} (ND18) = 58,7 μ M; IC_{50} (ND85) = 85,9 μ M; IC_{50} (kambinol) = 59 μ M)³¹¹. Uzimajući u obzir činjenicu da je svega nekoliko sub-mikromolarnih inhibitora SIRT2 otkriveno do danas, relativno jednostavan i inovativan hemijski skelet jedinjenja ND18 predstavlja odličnu polaznu tačku za buduće optimizacije u cilju poboljšavanja potentnosti ovog jedinjenja.



Slika 4.11. Rezultati eksperimentalne validacije SBVS. (A) Projekcije samoorganizujućih mapa hemijskog prostora za poznate inhibitore (levo) i kombinaciju poznatih inhibitora i virtuelno otkrivenih rezultata (desno). Bela zvezda označava poziciju dva novootkrivena SIRT2 inhibitora (ND18 i ND85) u hemijskom prostoru. (B) Rezultati in vitro evaluacije inhibicione aktivnosti i selektivnosti izabranih jedinjenja. Zvezdica (*) označava jedinjenje koje je interferiralo sa fluorometrijskim testom. (C) Fenotipske promene ćelija MDA-MB-231 uočene pri koncentraciji od 20 μ M ND18. (D) Distribucija faza ćelijskog ciklusa nakon 72 sata tretmana sa ND18 (M1 - sub G1, M2 - G0/G1, M3 - S, M4 - G2/M). (E) Nivoi E-kadherina nakon 72 sata tretmana sa koncentracijom od 20 μ M ND18.

4.1.7. Ipitivanje antitumorskog potencijala *hit* jedinjenja ND18

Antitumorska aktivnost ND18 je ispitana na ćelijskoj liniji trostruko negativnog adenokarcinoma dojke – MDA-MB-231. Ova ćelijska linija predstavlja visoko invazivnu, slabo diferenciranu ćelijsku liniju mezenhimalnog fenotipa i visoko aktiviranim signalnim putevima povezanim sa migracijom³¹². Nedavna studija koja je uporedila bazalni nivo ekspresije sirtuina u različitim ćelijskim linijama karcinoma, identifikovala je MDA-MB-231 kao ćelijsku liniju sa povećanom ekspresijom SIRT2³¹³. Sposobnost jedinjenja ND18 da inhibira proliferaciju MDA-MB-231 ćelija ispitana je u standardnom MTT testu. Smanjeno preživljavanje ćelija nakon 72 sata tretmana detektovano je u svim testiranim koncentracijama ND18 (Slika 4.12). S obzirom da je jedinjenje ND18 bilo umereno rastvorljivo u medijumu za gajenje ćelija, primećeno je zasićenje na tri najviše koncentracije (80, 40 i 20 μM) (Slika 4.12). Da bi se dalje ispitale antitumorske osobine ND18, ista ćelijska linija adenokarcinoma dojke je tretirana koncentracijom od 20 μM ND18 i pripremljena za apoptozni test i analizu ćelijskog ciklusa metodom protočne citometrije. Bojenje MDA-MB-231 ćelija sa Aneksinom V i 7-aminoktino-micin D (7-AAD) je pokazalo da ND18 nije izazvao apoptozu ili nekrozu pri testiranoj koncentraciji (Slika 4.12). Analiza ćelijskog ciklusa je pokazala da je ND18 uzrokovao zaustavljanje ćelijskog ciklusa u G2/M fazi nakon 72 sata tretmana (62,10% ćelija bilo je u G2 /M fazi u poređenju sa 20,15% ćelija tretiranih kontrolom) (Slika 4.11D).



Slika 4.12. Rezultati *in vitro* evaluacije jedinjenja ND18. A – Rezultati MTT eseja. B – Efekat 20 μM jedinjenja ND18 na apoptozu MDA-MD-231 ćelija nakon 72h (metoda dvokomponentnog bojenja sa Aneksin-V i 7-AAD).

Epitelijano-mezenhimalna tranzicija (EMT) predstavlja proces koji omogućava malignim ćelijama da potisnu epitelijske karakteristike i steknu mezenhimalni fenotip povezan sa povećanom pokretljivošću i invazivnošću. EMT predstavlja jedan od najvažnijih mehanizama povezanih sa inicijacijom tumora, metastazama i otpornošću na terapiju^{314,315}. Interesantno je da je nakon tretmana MDA-MB-231 ćelija sa 20 μM ND18 tokom 72 sata uočena značajna reverzija morfologije ćelija od mezenhimalne ka epitelijskoj (Slika 4.11C). Ova opservacija navela je na ispitivanje nivoa ekspresije E-kadherina, dobro poznatog epitelijskog markera EMT-a. Smanjenje ekspresije E-kadherina se smatra ključnim događajem u EMT-u kojim se povećava sklonost malignih ćelija ka invaziji i metastazama^{314,315}. Kako se može videti na Slici 4.11E, ND18 jedinjenje značajno stimuliše ekspresiju E-kadherina (80,41%) u poređenju sa kontrolom (1,13%), potvrđujući mikroskopijom uočenu reverziju epitelnih karakteristika u MDA-MD-231 ćelijskoj liniji. Ovo otkriće je u skladu sa ranijim studijama koje povezuju povećanje ekspresije E-kadherina sa inhibicijom SIRT2 u hepatocelularnom karcinomu, ili depleciju SIRT2 sa povećanom ekspresijom E-kadherina u trostrukom negativnom adenokarcinomu dojke^{170,316}. Osim toga, prethodno sprovedena istraživanja utvrdila su da je inhibicija SIRT2 povezana sa smanjenjem klinički relevantnih patoloških karakteristika trostrukog negativnog adenokarcinoma dojke, uključujući invazivne i migratorne osobine što je u skladu sa dobijenim rezultatima¹⁷⁰. Ukratko, novootkriveni selektivni inhibitor SIRT2 ND18 u koncentraciji od 20 μM se pokazao kao

necitotoksični inhibitor koji značajno indukuje zaustavljanje ćelijskog ciklusa u G2/M fazi i ometa EMT u MDA-MB-231 ćelijskoj liniji. Uzimajući u obzir inovativnost strukturnog skeleta i antitumorski potencijal, ND18 se može smatrati obećavajućim jedinjenjem za buduća istraživanja.

4.2. Razvoj *in silico* LBVS platforme za otkriće novih SIRT2 inhibitora

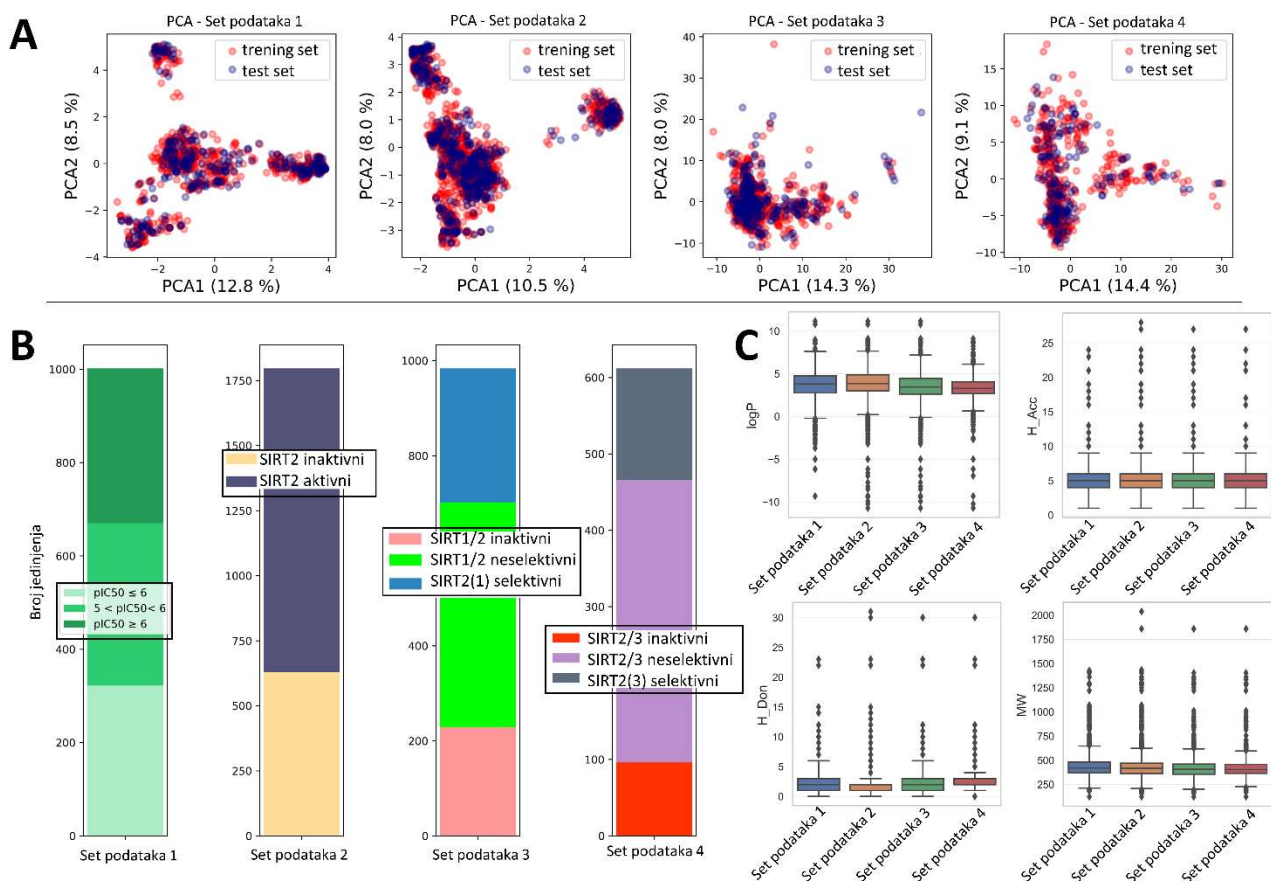
Sa prekliničkim studijama koje potvrđuju potencijal selektivnih SIRT2 inhibitora u tretmanu različitih vrsta malignih tumora, uz potvrđen sinergistički potencijal u kombinaciji sa poznatim antitumorskim terapijama, uključujući i imunoterapiju, selektivni inhibitori SIRT2 mogu predstavljati važnu grupu budućih terapeutika u nadolazećoj eri personalizovane terapije maligniteta (poglavlje 1.4.1). U cilju olakšavanja i ubrzavanja daljeg razvoja novih SIRT2 inhibitora, u okviru ovog dela istraživanja disertacije, kreirana je *in silico* platforma za brzi LBVS skrining i evaluaciju odnosa strukture i aktivnosti.

4.2.1. Opis seta podataka upotrebljenog za modelovanje odnosa strukture i aktivnosti SIRT2 inhibitora

Set podataka neophodan za uspostavljanje modela odnosa strukture i aktivnosti pripremljen je prikupljanjem svih dostupnih informacija iz ChEMBL baze podataka. Dodatni podaci su preuzeti pretragom repozitorijuma patenata. Pronađeno je ukupno 1797 jedinjenja sa zapisom o inhibitornoj aktivnosti na SIRT2. Ukupan broj jedinjenja je dalje podeljen na 4 seta podataka razmatrajući dostupne informacije o načinu izražavanja biološke aktivnosti kao i dodatnih zapisa o aktivnosti na dve izoforme sirtuina: SIRT1 i SIRT3. Set podataka 1 se sastojao od jedinjenja kod kojih je biološka aktivnost bila izražena kao pIC_{50} vrednost, set podataka 2 se sastojao od jedinjenja kod kojih je biološka aktivnost bila izražena i kao pIC_{50} i kao Inh%, dok su se setovi podataka 3 i 4 sastojali od jedinjenja koja su imala dodatne zapise i o aktivnosti na SIRT1 i SIRT3 (izražene kao pIC_{50} vrednosti, ili kao Inh%). Glavne karakteristike setova podataka su sumirane u Tabeli 4.3.

Tabela 4.3. Opis upotrebljenih setova podataka.

	Set podataka 1	Set podataka 2	Set podataka 3	Set podataka 4
Broj jedinjenja	1002	1797	984	612
Izraz aktivnosti	pIC_{50}	pIC_{50} and Inh%	Inh%	Inh%
Ciljna mesta	SIRT2	SIRT2	SIRT1, SIRT2	SIRT2, SIRT3
Način kodiranja aktivnosti u modelu	pIC_{50}	Aktivni, inaktivni	Selektivni, neselektivni, inaktivni	Selektivni, neselektivni, inaktivni



Slika 4.13. Deskriptivna statistička analiza setova podataka upotrebljenih za modelovanje. A – PCA analiza setova podataka; B – Distribucija pIC₅₀ vrednosti i klasa unutar setova podataka; C – Analiza ispunjenost Lipinski pravila unutar setova jedinjenja (H_Don – br. donora vodoničnih veza, H_Acc – br. akceptora vodoničnih veza).

Set podataka 1 se sastojao od 1002 jedinjenja u okviru kog je raspon pIC₅₀ vrednosti bio između 4 i 7,96. Set podataka 2, kao set sa kombinovanim podacima o biološkoj aktivnosti, je bio najbrojniji i sadržao je ukupno 1797 jedinjenja. Na osnovu kriterijuma predstavljenih u Materijalima i metodama, jedinjenja unutar seta 2 su bila klasifikovana na SIRT2 aktivna i SIRT2 inaktivna jedinjenja od kojih su SIRT2 aktivna jedinjenja predstavljala 2/3 ukupnog seta podataka. Setovi podataka 3 i 4 su na osnovu kriterijuma predstavljenih u Materijalima i metodama podeljeni u tri klase: selektivni inhibitori SIRT2, neselektivni SIRT1/2 odnosno SIRT2/3 inhibitori i klasu inaktivnih jedinjenja. Potrebno je naglasiti da su jedinjenja upotrebljena za formiranje setova podataka heterogenog porekla. Setovi podataka su kreirani uzimajući u obzir biološke aktivnosti dobijene različitim vidovima bioloških eseja (fluorimetrijski, luminiscentni, elektroforetski esej itd.) izvedenih pod različitim uslovima (različito vreme inkubacije, različite Km vrednosti upotrebljenih supstrata itd.). Naime, u cilju dobijanja modela širokog domena primenjivosti, potrebno je maksimalno povećati diverzitet seta podataka za generisanje modela. U slučaju QSAR modelovanja, idealan slučaj je da su svi biološki podaci za ispitivana jedinjenja generisati pod istim uslovima radi smanjenja varijabilnosti unutar seta podataka. Sa druge strane, za dobijanje robustnog modela širokog domena primenjivosti koji bi pokrio veliki deo hemijskog prostora i bio koristan u svrhe otkrića potpuno novih grupa inhibitora, potrebno je povećati strukturnu raznolikost jedinjenja za trening modela. U realnosti, veliki setovi raznovrsnih jedinjenja testiranih pod istim uslovima praktično da i ne postoje zbog čega je najčešće potrebno kombinovati podatke iz više različitih izvora. Kada je reč o regresionim modelima zasnovanim na pIC₅₀ vrednostima koji su i najosetljiviji na ovu varijabilnost, literaturni podaci ukazuju da kombinovanje podataka iz različitih izvora doprinosi samo sa 25% vrednosti standardne devijacije modela što je umeren nivo „šuma“ i prihvatljiv sa aspekta primene tih modela u svrhe otkrića novih kalasa inhibitora SIRT2.³¹⁷ Kada je

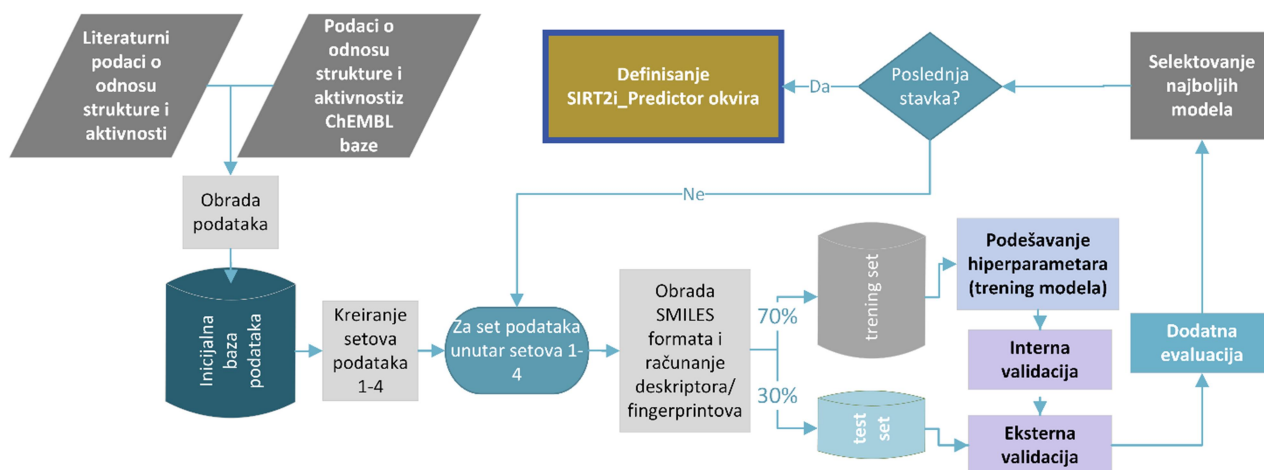
reč o klasifikacionim modelima, očekivana greška usled kombinovanja podataka iz različitih izvora je značajno smanjena pažljivim svrstavanjem jedinjenja u klase aktivnih, odnosno inaktivnih jedinjenja^{97,318–322}. Naime, umesto jasne granice između aktivnih i inaktivnih jedinjenja, kreiran je relativno širok interval aktivnosti nazvan „siva zona“ u kojoj je pretpostavljeno da jedinjenja ne mogu sa visokom pouzdanošću biti klasifikovana kao aktivna, ili inaktivna. Jedinjenja seta podataka koja su bila u navedenom intervalu su odbačena iz daljeg modelovanja pre formiranja finalnih setova podataka. Na ovaj način, obezbeđena je maksimalna nezavisnost klasifikacije od varijabilnosti odnosno šuma unutar seta podataka. Kriterijumi za definisanje klasa jedinjenja su detaljno predstavljeni u poglavlju 3.1.5.

Deskriptivnom statističkom analizom dobijenih setova podataka, zaključeno je da se najveći broj jedinjenja uklapa u Lipinski pravilo petice, uz svega nekoliko odstupanja, dok je balansirano podataka, odnosno distribucija različitih klasa i pIC_{50} vrednosti na zadovoljavajućem nivou (Slika 4.13). Pre formiranja modela, setovi podataka su podeljeni na set podataka za generisanje modela – trening set i set podataka za testiranje modela – test set primenom metode stratifikovanog randomizovanog uzorkovanja (30% test set, 70% trening set). Analiza glavnih komponenti ukazala je da je primenjena strategija efikasno uzorkovala jedinjenja test seta iz celokupnog hemijskog prostora pokrivenog inicijalnim setom podataka (Slika 4.13). Blagi disbalans u distribuciji tri klase unutar seta podataka 3 i 4 je rešen primenom SMOTE algoritma za dodatno uzorkovanje (eng. *oversampling*) manje zastupljenih klasa.

4.2.2. Razvoj i validacija modela mašinskog učenja

U okviru ovih istraživanja, kreirani su različiti regresioni, klasifikacioni binarni i multiklasifikacioni modeli mašinskog učenja. Za kreiranje modela upotrebjeno je pet različitih algoritama mašinskog učenja: RF, SVM, KNN, XGBoost, DNN i 4 različita deskriptora/fingerprinta za opis molekula – MACCS, ECFP4, ECFP6 i tradicionalni deskriptori iz Mordred seta. Zbog lakog računanja i velikog broja uspešnih heminformatičkih studija, ECFP i MACCS su u literaturi prepoznati kao jedni od najpouzdanijih i najčešće korišćenih fingerprintova. Generalni protokol za ovaj deo istraživanja je prikazan na Slici 4.14. Različiti regresioni i klasifikacioni modeli su trenirani kroz proces podešavanja hiperparametara primenom *Bayesian* optimizacije udružene sa petostrukom ukrštenom validacijom. U slučaju dubokih neuronskih mreža, zbog drugačije prirode samog algoritma, prilikom optimizacije u obzir su uzeti i parametri arhitekture mreže. Liste sa hiperparametrima razmatranih pri optimizaciji, kao i vrednosti optimizovanih hiperparametara za svaku kombinaciju algoritma mašinskog učenja i deskriptora/fingerprintova prikazana je u okviru Priloga 7.7.

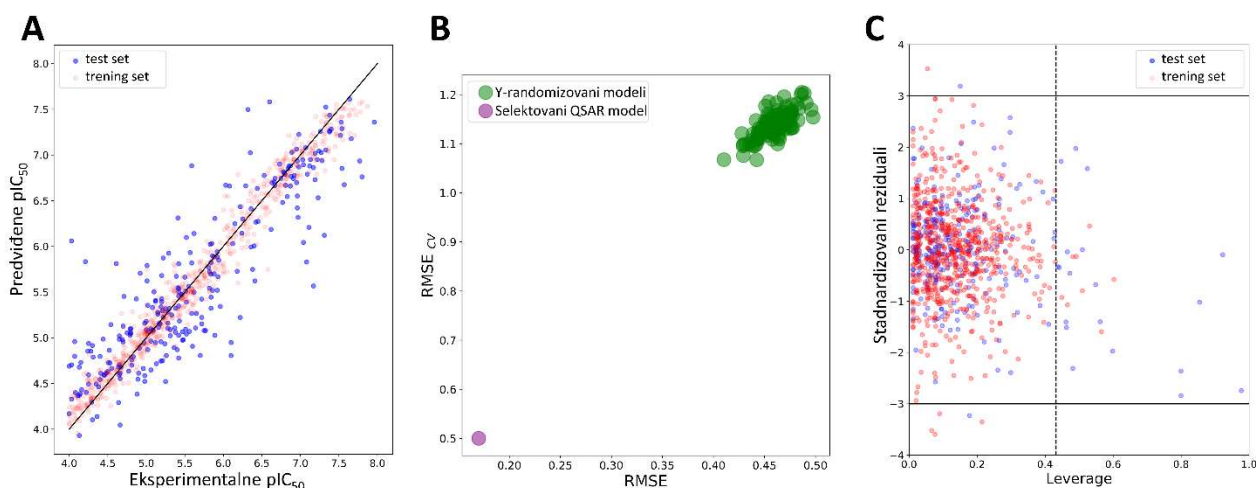
Prilikom treniranja modela, u obzir su uzeti parametri interne validacije, uključujući i parametre ukrštene validacije. Najbolji modeli su dalje reevalurani kroz eksternu validaciju primenom test seta. S obzirom na odgovarajuće nedeostatke parametara eksterne i interne validacije u proceni prediktivnosti modela na realnom uzorku, pri daljoj selekciji najboljih modela uzeti su u obzir i parametri dodatne validacije. Tip testa za dodatnu evaluaciju modela je odabran na osnovu tipa i svrhe svakog od modela što će biti detaljno opisano u daljem tekstu. Finalno odabrani modeli su upotrebljeni dalje za definisanje okvira *in silico* platforme za LBVS nazvane „*SIRT2i_Predictor*“.



Slika 4.14. Prikaz protokola za razvoj i validaciju modela i definisanje okvira *SIRT2i_Predictor*-a.

4.2.2.a. Regresioni model

Set podataka 1 je upotrebljen u svrhu formiranja regresionog modela. Procedurom za selekciju deskriptora (poglavlje 3.1.5), selektovano je ukupno 52 Mordred deskriptora za formiranje regresionih QSAR modela. Nasuprot deskriptorima, MACCS i ECFP4 fingerprintovi su upotrebljeni bez redukcije broja bitova. Nakon procesa optimizacije hiperparametara, kvalitet QSAR modela je praćen primenom sledećih parametara interne validacije: koeficijenta determinacije za trening set (R^2), korelacionog koeficijenta ukrštene validacije (Q^2), korena srednje kvadratne greške predviđanja (RMSE) trening seta ($RMSE_{int}$), i RMSE ukrštene validacije ($RMSE_{CV}$). Granične vrednosti na osnovu kojih su modeli selektovani za dalju validaciju su uzeti prema preporučenim literaturnim vrednostima ($R^2 > 0.6$, $Q^2 > 0.5$)³²³. Dodatno, metodom Y-randomizacije generisano je i evaluirano 100 QSAR modela nakon nasumičnog mešanja podataka unutar trening seta (Slika 4.15). Rezultati Y-randomizacije ukazali su da prediktivnost modela nije posledica slučajnosti čime je indikovana pouzdanost modela. U okviru svake kombinacije primenjenog algoritma mašinskog učenja i deskriptora/fingerprinta, zadržani su samo modeli koji su pokazali dobru internu moć predviđanja, stabilnost i robusnost.



Slika 4.15. Prikaz osnovnih karakteristika finalno selektovanog XGBoost:ECFP4 modela. A – Grafik odnosa eksperimentalnih i predviđenih pIC_{50} vrednosti; B – Rezultati Y-randomizacije; C – Domen primenjivosti modela.

Tabela 4.4. Prikaz parametara eksterne validacije treniranih regresionih modela.

Algoritam	Molekulska osobina	R_{ext}^2	$RMSE_{ext}$	\bar{r}_m^2	Δr_m^2	Q_{F1}^2	Q_{F2}^2	Q_{F3}^2	CCC
RF	Deskriptor	0,7	0,55	0,52	0,27	0,7	0,7	0,7	0.81
	ECFP4	0,75	0,5	0,6	0,23	0,75	0,75	0,75	0.85 ^a
	MACCS	0,71	0,53	0,55	0,26	0,71	0,71	0,71	0.82
	ECFP6	0,77	0,48	0,62	0,21	0,77	0,77	0,76	0.86 ^a
SVR	Deskriptor	0,62	0,61	0,44	0,31	0,62	0,62	0,62	0.77
	ECFP4	0,74	0,51	0,63	0,13 *	0,74	0,74	0,73	0.84
	MACCS	0,68	0,57	0,55	0,21	0,68	0,68	0,68	0.81
	ECFP6	0,74	0,51	0,63	0,18 *	0,74	0,74	0,74	0.86 ^a
XGBoost	Deskriptor	0,67	0,58	0,53	0,25	0,68	0,68	0,68	0.82
	ECFP4	0,75	0,5	0,64	0,17	0,74	0,74	0,74	0.86
	MACCS	(0,79) ^b	(0,46) ^b	(0,7) ^b	(0,17) ^{*b}	(0,75) ^b	(0,75) ^b	(0,74) ^b	(0.86) ^{a,b}
	ECFP6	0,71	0,53	0,58	0,24	0,7	0,7	0,7	0.82
KNN	Deskriptor	0,68	0,56	0,56	0,23	0,68	0,68	0,68	0.86 ^a
	ECFP4	0,74	0,51	0,64	0,13 *	0,74	0,74	0,74	0.87 ^a
	MACCS	0,6	0,63	0,47	0,16 *	0,6	0,6	0,6	0.79
	ECFP6	0,76	0,49	0,66	0,12	0,76	0,76	0,76	0.87
DNN	Deskriptor	(0,77) ^b	(0,48) ^b	(0,68) ^b	(0,11) ^{*b}	(0,76) ^b	(0,76) ^b	(0,76) ^b	(0.87) ^{a,b}
	ECFP4	0,66	0,58	0,57	0,03 *	0,66	0,66	0,66	0.81
	MACCS	0,74	0,51	0,63	0,18 *	0,73	0,73	0,73	0.84
	ECFP6	0,68	0,56	0,56	0,16 *	0,68	0,68	0,67	0.80
Kriterijum	0,73	0,52	0,63	0,17 *	0,73	0,73	0,73	0.81	
		>0,6		>0,5	<0,2	>0,7	>0,7	>0,7	>0,85

* Vrednosti unutar zadatog kriterijuma $\Delta r_m^2 < 0,2$; ^a Vrednosti iznad kriterijuma $CCC > 0,85$; ^b Vrednosti unutar zagrada predstavljaju vrednosti nakon korekcije domena primenjivosti.

Eksterna validacija izvedena je u cilju finalne potvrde moći predviđanja modela van prostora definisanog trening setom. U skladu sa OECD principima (eng. Organization for Economic Co-operation and Development – OECD)⁹⁶, eksterna validacija predstavlja ključni dokaz kvaliteta QSAR modela i može se izvesti praćenjem koeficijenta determinacije R_{ext}^2 koji bi trebalo da bude veći od 0,6 i korena srednje kvadratne greške $RMSE_{ext}$ koji bi trebalo da bude što niži. Modeli formirani za svaku kombinaciju upotrebljenog algoritma i deksriptroa/fingerprinta su pokazali ekvivalentnu prediktivnost na test setu uz blagu prednost modela formiranih upotrebom ECFP4 i ECFP6 fingerprinta. Ipak, literaturni podaci ukazuju na to da oslanjanje na samo R^2 eksterne validacije može rezultovati u preuranjenoj proceni prediktivnosti modela. Naime, R^2 umnogome zavisi od raspona vrednosti zavisne varijable unutar test seta i distribucije vrednosti u odnosu na srednju vrednost trening seta/test seta. Zbog toga, u cilju kvalitetnije validacije regresionih QSAR modela mnogi autori preporučuju primenu i dodatnih parametara koji se, za razliku od jednostavnog

parametra poput R^2 , oslanjaju i na procenu preciznost modela, tačnosti modela, i eliminišu pristrasnost uvedenu rasponom vrednosti zavisne varijable. Dodatni parametri koji su upotrebljeni u validaciji regresionih modela su: \bar{r}_m^2 metrika predložena od strane Roy i saradnika ($\bar{r}_m^2 > 0.5$ i $\Delta r_m^2 < 0.2$)^{101,102}, zatim Q_{Fn}^2 metrika i CCC metrika sa kriterijumima predloženim od strane autora Chirico i Gramatica^{103,324} (Q_{F1}^2 , Q_{F2}^2 , and $Q_{F3}^2 > 0.7$, i $CCC > 0.85$) (Tabela 4.4). Dodatna analiza je izvedena i primenom kriterijuma predloženih od strane autora Golbrikh i Tropsha³²³ ($(R^2 - R_0^2) / R^2 < 0.1$ ili $(R^2 - R_0'^2) / R^2 < 0.1$, $0.85 \leq k$ (ili k') ≤ 1.15 i $|R^2 - R_0'^2| < 0.3$) (Tabela 4.4 i Prilog 7.8). Svi kreirani modeli su zadovoljili većinu kriterijuma, osim Δr_m^2 i CCC, zbog čega je na osnovu ova dva kriterijuma izvršena finalna selekcija modela. Jedina dva modela koja su zadovoljila sve navedene kriterijume bila su XGBoost:ECFP4 model i KNN:ECFP6 model, pri čemu je model KNN:ECFP6 pokazao manju robustnost u procesu interne validacije (Tabela 4.4).

Kao jedan od osnovnih OECD zahteva prilikom definisanja novih QSAR modela predstavlja i definisanje domena primenjivosti modela. Jedna od najčešće primenjivanih metoda pri definisanju QSAR modela jeste tzv. *leverage* metoda. Vrednosti *leverage*-a su proporcionalne *Mahalanobis*-ovom rastojanju navedenog jedinjenja u odnosu na centroid trening seta. Dobijene vrednosti *leverage*-a za jedinjenja su predstavljena kroz *Williams*-ov grafik koji prikazuje odnos *leverage* vrednosti svakog jedinjenja i standardizovanih reziduala (Slika 4.15). Domen primenjivosti je izračunat i upoređen za dva najbolja modela, izdvojena nakon eksterne validacije. Granice domena primenjivosti su postavljene na osnovu literaturnih preporuka i način računanja je predstavljen detaljnije u poglavlju 3.1.5. KNN:ECFP6 model je pokazao uži domen primenjivosti sa 39 jedinjenja test seta izvan definisanih granica domena primenjivosti. U slučaju XGBoost:ECFP4 modela, 24 jedinjenja su bila van definisanih granica domena primenjivosti. Očekivano, izbacivanjem jedinjenja koja su bila van definisanog domena primenjivosti modela došlo je do značajnih poboljšanja u vrednostima parametara eksterne validacije pri čemu je značajnije poboljšanje zabeleženo kod XGBoost:ECFP4 modela (Tabela 4.4). U zaključku, rezultati validacije modela su ukazali na lošiju moć predviđanja i pokrivenost hemijskog prostora u slučaju KNN:ECFP6 modela. Uzimajući u obzir da je za globalne QSAR modele od ključnog značaja postići balans između pokrivenosti hemijskog prostora, robusnosti i moći predviđanja modela, XGBoost:ECFP4 je selektovan kao finalni model za dalji rad. Ipak, treba spomeniti i ograničenja samog regresionog modela koja su posledica limitiranosti seta podataka upotrebljenog za generisanje modela (set podataka 1). Pored standardnog problema sa varijabilnošću podataka unutar velikih baza podataka usled heterogenosti u metodama i uslovima određivanja IC_{50} vrednosti (diskutovano iznad), opseg aktivnosti unutar seta podataka ponekad takođe može predstavljati ograničavajući faktor za praktičnu primenu dobijenih modela. Posmatrajući distribuciju aktivnosti unutar seta podataka 1, može se uočiti da je najveće broj jedinjenja pripadao potentnim i umereno potentnim SIRT2 inhibitorima (914 jedinjenja nasuprot 88 jedinjenja koja se mogu okarakterisati kao neaktivna, ili slabo aktivna). U tom kontekstu može se govoriti i o ograničenjima primene datog modela, gde bi upotreba modela u svrhe virtuelnog skrininga bila manje pouzdana u odnosu na primenu u optimzaciji aktivnosti umereno potentnih jedinjenja.

4.2.2.b. Binarni klasifikacioni model

S obzirom na ograničenu primenjivost regresionog modela u VS studijama, u nastavku studije kreiranu su binarni klasifikacioni QSAR modeli. Posmatranje jedinjenja kroz klase, umesto kroz IC_{50} vrednosti, značajno doprinosi smanjenju greške unete usled heterogenosti u izvorima podataka. Dodatno, osim primenom IC_{50} vrednosti, jedinjenja se mogu klasifikovati na aktivna i inaktivna upotrebom i procenta inhibicije. Procenat inhibicije je znatno češći podatak unutar ChEMBL baze podataka pre svega jer ne zahteva opsežnu studiju odnosa doze i odgovora, već se brojeći izraz inhibitorne aktivnosti dobija pri merenju aktivnosti biološkog ciljnog mesta od interesa pri jednoj definisanoj koncentraciji inhibitora. Posledično, set podataka kreiran u svrhe formiranja klasifikacionog modela (set podataka 2) sadržao je veći broj jedinjenja, pokazao je veću hemijsku

raznolikost i balasniran odnos aktivnih i inaktivnih predstavnika. Pravila na osnovu kojih su jedinjenja podeljena u klase predstavljena su u poglavlju 3.1.5.

Shodno opštem protokolu studije (Slika 4.14), binarni klasifikacioni modeli formirani su primenom pet različitih algoritama mašinskog učenja i četiri različita molekulska prikaza kroz proces *Bayes*-ove optimizacije hiperparametara. Pre formiranja binarnih klasifikacionih modela, inicijalni set Mordred deskriptora je sveden na 233 deskriptora kroz proces selekcije najznačajnijih deskriptora. Fingerprintovi su korišćeni bez dalje redukcije broja bitova. Parametri interne validacije (Prilog 7.8) ukazali su na dobru internu prediktivnost, stabilnost i robusnost svakog od optimizovanih modela. Eksterna prediktivnost modela ispitana je primenom test seta kroz praćenje balansirane tačnosti (eng. *Balanced Accuracy* – BA), opoziva (eng. *Recall*), preciznosti, F1-skora, *Matthews*-ov korelacioni koeficijenta (MCC) i površine ispod ROC (eng. *Receiver Operating Characteristics* – ROC) krive (ROC_AUC). Gotovo svi modeli pokazali su odličnu eksternu prediktivnost, dok su se RF, SVC i DNN modeli formirani primenom Mordred deskriptora pokazali kao najbolji (Tabela 4.5).

U praktičnim uslovima primene modela za potrebe virtuelnog skrininga (VS), jedan od bitnih aspekata procene kvaliteta VS modela je i sposobnost modela da unutar seta izgrađenog od velikog broja neaktivnih i malog broja aktivnih jedinjenja da daje prioritet aktivnim jedinjenjima. U cilju dalje evaluacije praktične primene kreiranih VS modela, generisan je set podataka sastavljen od molekula mamaca koji je sadržao oko 20000 virtuelno generisanih neaktivnih jedinjenja. Virtuelni mamci su generisani primenom principa forsiranjem 2D strukturne različitosti uz očuvanje fizičko-hemijskih sličnosti u odnosu na aktivna jedinjenja. Spajanjem seta mamaca sa test setom, dobijen je odnos aktivnih:inaktivnih = 1:40 što predstavlja realniji scenario u primeni VS modela. Evaluacijom sposobnosti generisanih modela da prioritizuju aktivna jedinjenja iz seta molekula mamaca, izdvojeni su najbolji modeli. Pored već spomenutih parametara eksterne validacije, praćeni su i tzv. parametri metrike obogaćenja. Faktori ranog obogaćenja (EF 0.5%, 1%, 2%, i 5%), izraženi kao površina ispod ROC krive nakon pronalaska 0.5, 1, 2 ili 5% lažno pozitivnih jedinjenja, predstavljaju jedan od najčešće korišćenih parametara pri proceni kvaliteta modela za VS (Tabela 4.6).

Tabela 4.5. Parametri eksterne validacije binarnih modela.

Algoritam	Molekulska osobina	BA	MCC	ROC_AUC	Preciznost ^a	Opoziv ^a	F1 ^a
RF	Deskriptor	0,88	0,74	0,94	0,86	0,88	0,87
	ECFP4	0,84	0,66	0,92	0,82	0,84	0,83
	MACCS	0,82	0,62	0,91	0,8	0,82	0,81
	ECFP6	0,85	0,68	0,92	0,83	0,85	0,84
SVR	Deskriptor	0,88	0,74	0,95	0,87	0,88	0,87
	ECFP4	0,81	0,63	0,9	0,82	0,81	0,82
	MACCS	0,8	0,59	0,87	0,79	0,8	0,79
	ECFP6	0,79	0,62	0,9	0,83	0,79	0,81
XGBoost	Deskriptor	0,86	0,72	0,94	0,85	0,86	0,85
	ECFP4	0,81	0,62	0,91	0,8	0,81	0,81
	MACCS	0,8	0,6	0,9	0,8	0,8	0,8
	ECFP6	0,81	0,62	0,91	0,81	0,81	0,81
KNN	Deskriptor	0,79	0,56	0,88	0,77	0,79	0,77
	ECFP4	0,82	0,62	0,9	0,8	0,82	0,81
	MACCS	0,82	0,62	0,88	0,8	0,82	0,81
	ECFP6	0,84	0,65	0,91	0,81	0,84	0,82
DNN	Deskriptor	0,89	0,75	0,94	0,85	0,86	0,86
	ECFP4	0,83	0,65	0,91	0,8	0,81	0,8
	MACCS	0,8	0,58	0,89	0,8	0,8	0,8
	ECFP6	0,82	0,64	0,9	0,79	0,82	0,8

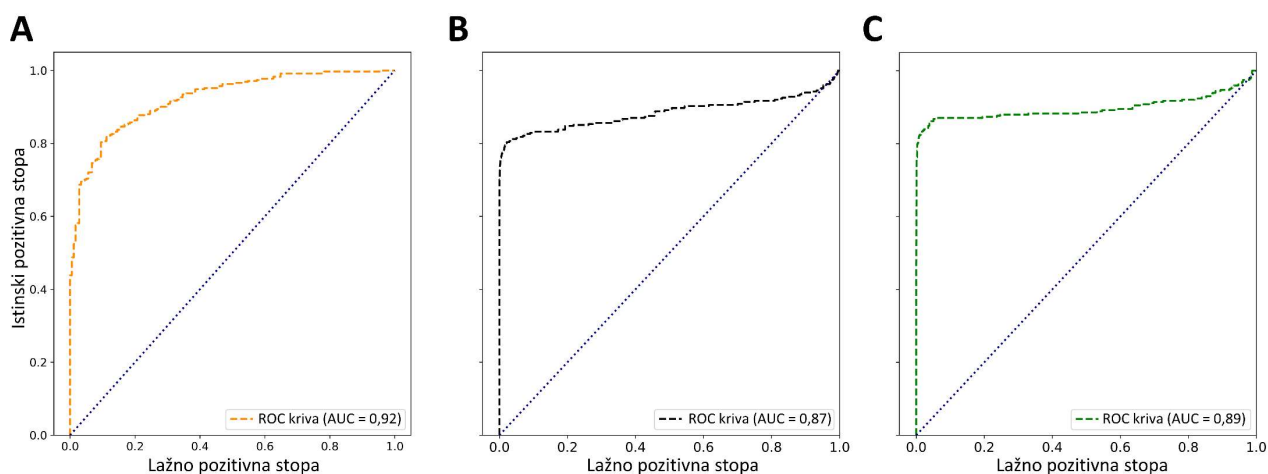
^a Srednje vrednosti parametara unutar klasa

Tabela 4.6. Parametri validacije binarnog modela primenom seta molekula mamaca.

Algoritam	Molekulska osobina	BA	MCC	ROC_AUC	Preciznost ^a	Opoziv ^a	F1 ^a	EF _{05%}	EF _{1%}	EF _{2%}	EF _{5%}
RF	Deskriptor	0,68	0,09	0,87	0,51	0,68	0,35	0,63	0,67	0,68	0,73
	ECFP4	0,81 (0,9) ^b	0,19 (0,52) ^b	0,87 (0,89) ^b	0,53 (0,67) ^b	0,81 (0,9) ^b	0,49 (0,73) ^b	0,74 (0,74) ^b	0,74 (0,74) ^b	0,76 (0,76) ^b	0,77 (0,8) ^b
	MACCS	0,66	0,08	0,82	0,51	0,66	0,35	0,55	0,56	0,59	0,62
	ECFP6	0,75	0,14	0,87	0,52	0,75	0,43	0,72	0,74	0,76	0,78
SVR	Deskriptor	0,69	0,1	0,89	0,51	0,69	0,36	0,43	0,56	0,62	0,71
	ECFP4	0,46	-0,06	0,8	0,48	0,46	0,05	0,75	0,75	0,75	0,76
	MACCS	0,62	0,06	0,83	0,51	0,62	0,32	0,39	0,61	0,68	0,74
	ECFP6	0,47	-0,07	0,8	0,47	0,47	0,03	0,76	0,76	0,77	0,77
XGBoost	Deskriptor	0,71	0,11	0,85	0,51	0,71	0,39	0,41	0,44	0,48	0,54
	ECFP4	0,74	0,13	0,87	0,52	0,74	0,42	0,35	0,39	0,43	0,52
	MACCS	0,64	0,07	0,73	0,51	0,64	0,32	0	0	0,02	0,2
	ECFP6	0,71	0,11	0,85	0,51	0,71	0,39	0,37	0,38	0,44	0,5
KNN	Deskriptor	0,66	0,08	0,76	0,51	0,66	0,37	0,09	0,23	0,26	0,29
	ECFP4	0,72	0,12	0,8	0,52	0,72	0,41	0	0	0	0
	MACCS	0,64	0,07	0,75	0,51	0,64	0,33	0	0	0	0
	ECFP6	0,72	0,11	0,8	0,52	0,72	0,41	0	0	0	0
DNN	Deskriptor	0,72	0,12	0,8	0,51	0,71	0,38	0	0	0	0
	ECFP4	0,73	0,13	0,84	0,52	0,73	0,43	0,1	0,25	0,32	0,41
	MACCS	0,69	0,1	0,79	0,51	0,62	0,29	0,04	0,08	0,17	0,23
	ECFP6	0,67	0,09	0,81	0,51	0,67	0,38	0,17	0,25	0,34	0,43

^a Srednje vrednosti parametara unutar klasa; ^b Vrednosti unutar zagrada predstavljaju vrednosti nakon korekcije domena primenljivosti.

Nakon ponovne evaluacije binarnih modela primenom neuravnoteženog seta mamaca, RF:ECFP4 binarni model je pokazao najbolju moć predviđanja, visoku senzitivnost, specifičnost, preciznost, robusnost i mogućnost ranog obogaćenja. Nakon samo 0,5% lažno pozitivnih jedinjenja, RF:ECFP4 binarni model je uspeo da prepozna 70% aktivnih jedinjenja što ga čini odličnim modelom za virtuelni skrining. Potrebno je istaći i činjenicu da je set mamaca bio sastavljen od strukturno različitih molekula čime je RF:ECFP4 binarni model demonstrirao mogućnost funkcionisanja u širem hemijskom prostoru u odnosu na set podataka koji je korišćen za generisanje modela.



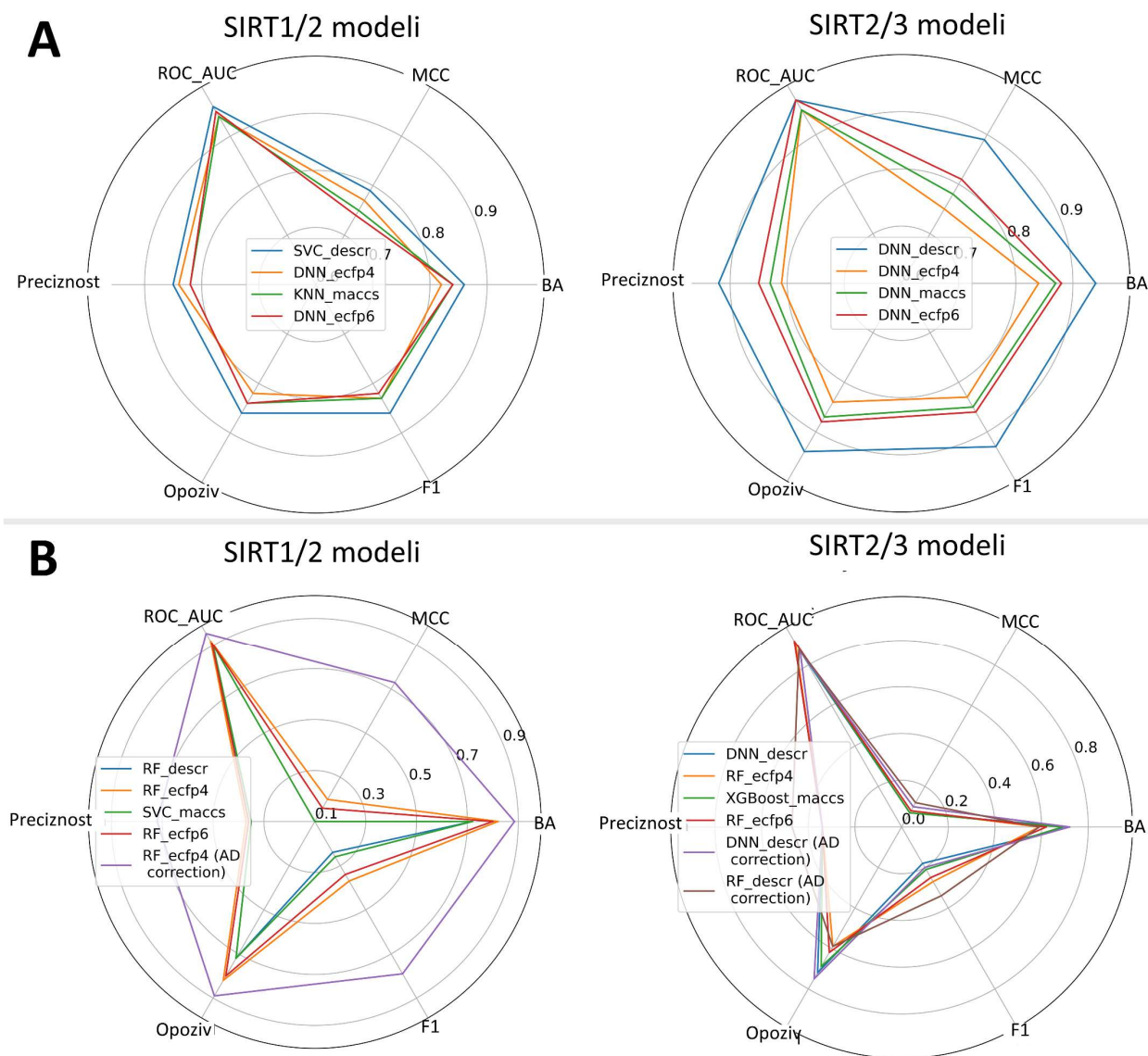
Slika 4.16. Prikaz performansi finalnog binarnog RF:ECFP4 modela. A – Rezultati eksterne validacije; B – Rezultati validacije na setu molekula mamaca; C – Rezultati na setu molekula mamaca nakon korekcije domena primenljivosti.

Domen primenljivosti binarnih modela definisan je primenom metode neodređene zone (eng. *indeterminate-zone*). Zona neodređenosti je definisana kroz verovatnoće predviđanja od $0,5 \pm 0,1$. Na osnovu ove metode, verovatnoće predviđanja koje su u opsegu definisanim zonom neodređenosti smatraju se nepouzdanim i obrnuto. Računajući statističke parametre modela samo u okviru jedinjenja unutar domena primenljivosti, zabeležena su značajna poboljšanja modela RF:ECFP4 čime je ovaj model selektovan kao finalni (Tabela 4.6, Slika 4.16).

4.2.2.c. Višeklasni klasifikacioni modeli

SIRT2 kao predstavnik klase I sirtuina deli veliku strukturnu sličnost sa ostalim predstavnicima svoje klase. SIRT1 i SIRT3 predstavljaju strukturno najpribližnije homologe SIRT2. Usled sličnosti u strukturi na nivou vezivnog mesta inhibitora, postizanje selektivne inhibicije SIRT2 predstavlja kompleksan zadatak. Najnoviji rezultati istraživanja iz oblasti hemijske biologije sirtuina ukazali su na vezu između selektivnosti i bezbednosti SIRT2 inhibitora zbog čega se SIRT2 selektivnost sve češće postavlja kao jedan od imperativa razvoja ove grupe potencijalnih terapeutika. Značajan deo jedinjenja iz ChEMBL baze sa SIRT2 inhibitornom aktivnošću poseduje i zapise o selektivnosti prema SIRT1 i SIRT3. Ove informacije su pretočene u setove podataka 3 i 4 koji su dalje upotrebljeni za treniranje višeklasnih klasifikacionih modela. Kao što je objašnjeno u poglavlju 4.2.1, jedinjenja unutar setova 3 i 4 su podeljena u tri klase (selektivni inhibitori, neselektivni inhibitori neaktivna jedinjenja) u cilju uspostavljanja višeklasnih klasifikacionih modela. Ipak, setovi 3 i 4 pored manjeg broja jedinjenja u odnosu na setove 1 i 2, imaju i problem sa većim disbalansom unutar klasa što može uticati na moć predviđanja klasifikacionih modela (Slika 4.13).

U skladu sa generalnim protokolom studije, grupa modela za procenu selektivnosti izgrađena je testiranjem kombinacija svih prethodno spomenutih algoritama mašinskog učenja i molekularnih reprezentacija. Set podataka 3 je upotrebljen za treniranje modela SIRT1/2 selektivnosti dok je set podataka 4 upotrebljen za treniranje modela SIRT2/3 selektivnosti. Usled malog broja podataka i velikog disbalansa seta jedinjenja koji su imali zapise aktivnosti za sve tri sirtuinske izoforme, grupni model SIRT1/2/3 nije uspostavljen. Nakon procedure selekcije najznačajnijih deskriptora, 270 Mordred deskriptora je selektovano za treniranje SIRT1/2 modela, dok je 316 Mordred deskriptora selektovano za treniranje SIRT2/3 modela. Fingerprintovi su i ovaj put upotrebljeni bez dalje redukcije broja bitova. Višeklasni klasifikacioni modeli SIRT1/3 i SIRT2/3 izgrađeni su i evaluirani na sličan način poput binarnih klasifikacionih modela. Parametri interne validacije ukazali su na dobru internu prediktivnost treniranih modela optimizovanih hiperparametara dok su parametri eksterne validacije ukazali na predominantnost modela uspostavljenih primenom Mordred deskriptora i RF, DNN ili SVC algoritama (Slika 4.17 i Prilog 7.8). Interesantno je uočiti da su u slučaju modela treniranih na najmanjem setu podataka (set podataka 4), DNN SIRT2/3 modeli postigli najbolji rezultat što potvrđuje sposobnost dubokih neuronskih mreža da efikasnije uče na ograničenom setu podataka (Slika 4.17).



Slika 4.17. Prikaz statističkih parametara multiklasifikacionih modela. A – Rezultati eksterne validacije na test setu; B – Rezultati validacije na setu molekula mamaca.

Slično kao i kod binarnih modela, praktična primena ove vrste klasifikacionih modela takođe može obuhvatati i virtuelni skrining. Treba imati na umu da su setovi 4 i 5 sadržali manji broj inaktivnih jedinjenja u odnosu na set podataka 2 što značajno ograničava primenu modela selektivnosti. U cilju evaluacije primenjivosti ovih modela u svrhe virtuelnog skrininga, kreiran je set mamaca (aktivni:inaktivni = 1:40). U uslovima velikog opterećenja neaktivnim jedinjenjima prisutnog u setu mamaca, istakli su se modeli formirani na osnovu ECFP4 fingerprinta i RF algoritma, uz izuzetak DNN:deskriptorskog SIRT2/3 modela koji je pokazao statistiku uporedivu sa RF:ECFP4 SIRT2/3 modelom (Slika 4.17 i Prilog 7.8). Treba istaći da su SIRT2/3 modeli trenirani na najmanjem setu podataka zbog čega su svi imali lošije statističke parametre u odnosu na SIRT1/2 modele i njihova primena bi trebalo da bude svedena na jedinjenja koja su topološki slična jedinjenjima trening seta.

U cilju definisanja jasnijih okvira za primenu navedenih modela, definisani su i domeni primenjivosti sličnom metodom kao kod binarnih modela. S obzirom da su višeklasni klasifikacioni modeli sadržali tri klase, limiti klasifikacione verovatnoće su postavljeni na $> 0,5$. Preračunavanjem statističkih parametara samo za jedinjenja koja su pripadala okvirima domena primenjivosti (6000 jedinjenja), dobijena su velika poboljšanja u statističkim parametrima SIRT1/2 modela gde se kao najbolji istakao SIRT1/2 RF:ECFP4 model (Slika 4.17 i Prilog 7.8). Nasuprot SIRT1/2 modelima, dva istaknuta SIRT2/3 modela (RF:ECFP4 i DNN:deskriptorski SIRT2/3 model) ispoljila su vrlo

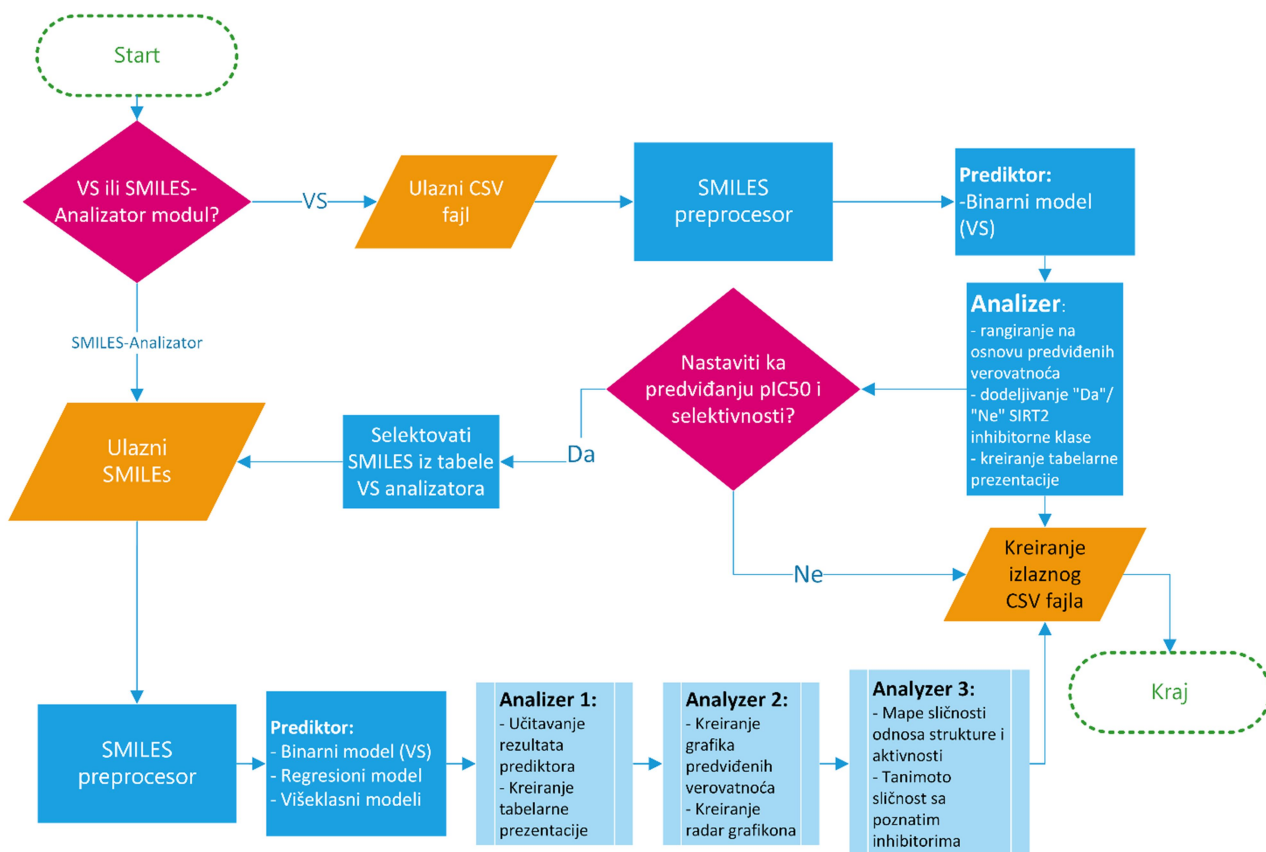
umerena poboljšanja statističkih parametara unutar definisanih domena primenjivosti s tom razlikom da je DNN:deskriptorski model unutar svog domena primenjivosti sadržao 19000 jedinjenja dok je RF:ECFP4 model sadržao svega 9000 jedinjenja. Ovi rezultati ukazali su na značajno veću pokrivenost hemijskog prostora primenom DNN:deskriptorskog SIRT2/3 modela zbog čega je on selektovan za dalju primenu.

Treba napomenuti da su modeli selektivnosti trenirani na setovima sa manjim brojem i diverzitetom jedinjenja i da njihova sposobnost prepoznavanja aktivnih jedinjenja ostaje niža u odnosu na binarni model. Uzimajući u obzir navedeno, modeli selektivnosti su najprikladniji za dodatnu evaluaciju rezultata virtuelnog skrininga dobijenih nakon primene binarnog modela, odnosno jedinjenja koja su prvobitno predviđena kao aktivna primenom binarnog modela dok bi sve oprečne predikcije trebalo uzeti sa dozom opreza.

4.2.3. Definisanje okvira *SIRT2i_Predictor* platforme za otkriće novih SIRT2 inhibitora

U cilju povećanja dostupnosti kreiranih modela i obezbeđivanja dobre prakse u njihovoj primeni, okvir za primenu navedenih modela je definisan i kodiran kreiranjem Python aplikacije nazvane *SIRT2i_Predictor*. Protokol *SIRT2i_Predictor*-a se sastoji od sledećih celina: 1) selektor modula; 2) SMILES preprocesor; 3) prediktori i 4) analizatori (Slika 4.18). U cilju efikasnije diseminacije *SIRT2i_Predictor*-a široj javnosti, kreiran je i grafički interfejs kojim se značajno olakšava i pospešuje upotreba navedene aplikacije (Slika 4.19).

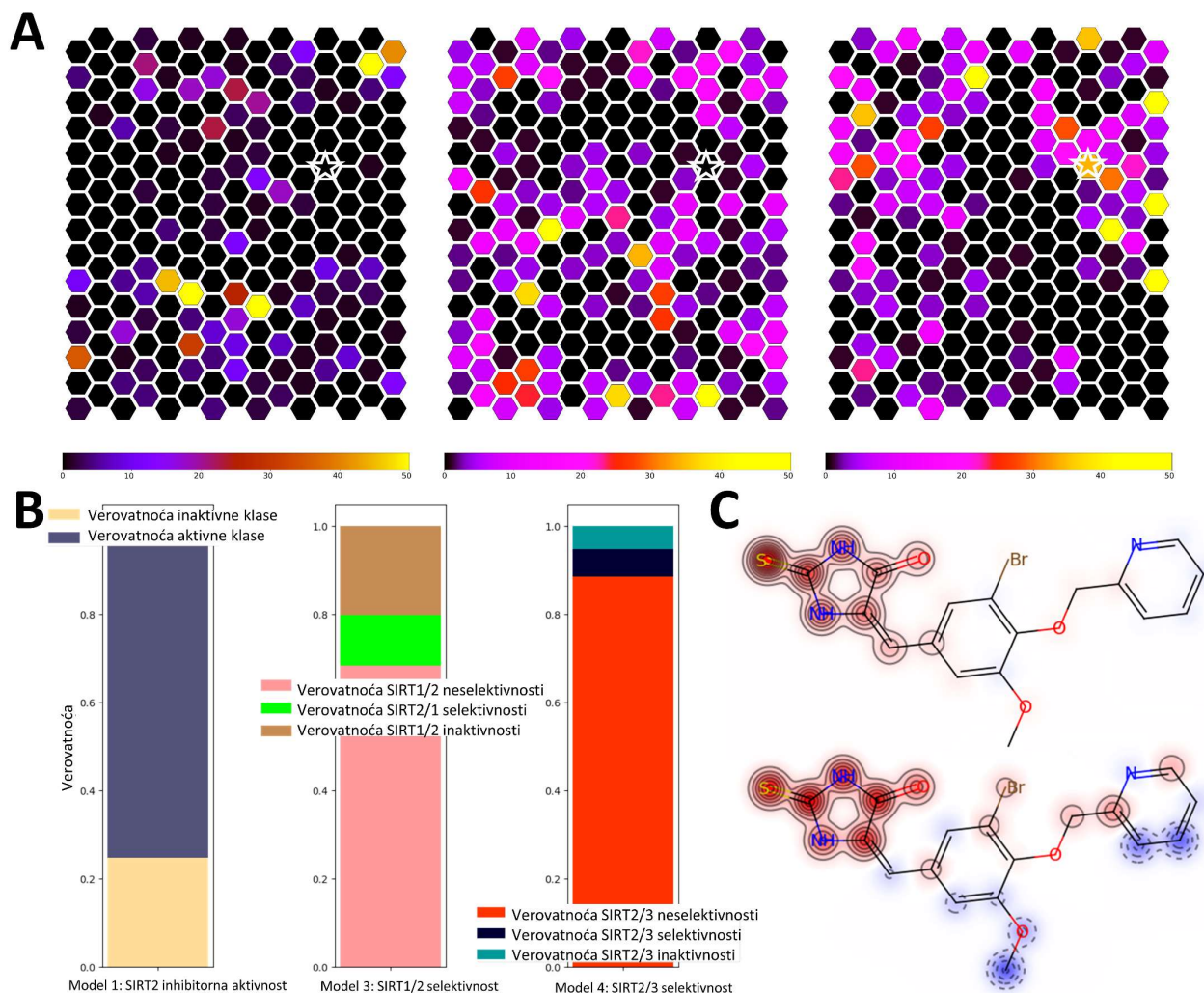
Selektor modula dozvoljava korisnicima odabir dva modula – virtuelni skrining (VS) i SMILES analizador. U okviru VS modula, *SIRT2i_Predictor* kao ulaznu komponentu koristi CSV fajl sa listom jedinjenja enkodiranih u obliku SMILES formata sa ili bez predefinisanih indentifikacionih oznaka. Maksimalna veličina CSV fajla iznosi 200MB i nakon unosa jedinjenja, *SIRT2i_Predictor* automatski formatira jedinjenja primenom SMILES-preprocesora i priprema ih za predviđanje aktivnosti, odnosno selektivnosti. U skladu sa prethodno diskutovanim ograničenjima regresionog modela i modela selektivnosti u uslovima virtuelnog skrininga gde se očekuje veliki broj inaktivnih jedinjenja, u prediktor VS modula uključen je samo binarni RF:ECFP4 model. Isključivanjem ostalih modela iz VS modula aplikacije ostvarena je velika ušteda vremena pri predviđanju aktivnosti nepoznatih jedinjenja što ima veliki praktični značaj u skriningu velikih baza podataka. Kao izlaznu informaciju, VS modul *SIRT2i_Predictor*-a daje CSV fajl sa listom jedinjenja, predviđenim verovatnoćama binarnog modela i dodeljenim klasama (SIRT2 aktivno jedinjenje i SIRT2 neaktivno jedinjenje). Grafički interfejs *SIRT2i_Predictor*-a omogućava i vizuelizaciju CSV fajla unutar samog programa. Navedene funkcionalnost VS modula čine *SIRT2i_Predictor* odličnim alatom za prioritizaciju jedinjenja pri *in vitro* visoko propusnom skriningu u cilju uštede vremena i resursa.



Slika 4.18. Definisani okvir rada aplikacije *SIRT2i_Predictor*.

Preostali modeli (regresioni i modeli selektivnosti) implementirani su u okviru drugog modula – SMILES analizatora koji je namenjen detaljnijoj analizi rezultata virtuelnog skrininga. Ovaj modul kao ulaznu informaciju koristi SMILES notacije struktura jedinjenja koja se direktno kopiraju na predviđeno mesto u grafičkom interfejsu. SMILES analizator koristi SMILES preprocesor za pripremu jedinjenja za predviđanje aktivnosti. U okviru prediktora SMILES analizatora, nalaze se sva 4 razvijena modela mašinskog učenja. Slično VS modulu, i SMILES analizator može da podrži rad sa većim brojem jedinjenja, ali zbog pripreme jedinjenja za svaki model ponaosob, rad sa velikim brojem jedinjenja je znatno sporiji u odnosu na VS modul. Kao izlaznu informaciju SMILES analizator nudi detaljne tekstualne, numeričke i grafičke rezultate predviđanja svakog od modela. U slučaju regresionog modela, rezultati uključuju tabelarni prikaz pIC_{50} vrednosti dok u slučaju klasifikacionih modela rezultati uključuju tabelarni prikaz verovatnoće predviđanja kao i nazive dodeljenih klasa. Grafički prikazi uključuju radar grafikone predviđanja svakog od 4 modela, histograme sa verovatnoćama predviđanja klasifikacionih modela za procenu domena primenjivosti klasifikacije, kao i *leverage* grafik domena primenjivosti regresionog modela. U cilju omogućavanje detaljnije analize odnosa strukture i aktivnosti i uspostavljanja hipoteza o daljim hemijskim modifikacijama unutar studija optimizacije *lead* struktura, u okviru *SIRT2i_Predictor*-a uvedena je i funkcionalnost računanja mapa doprinosa svakog atoma rezultatu predviđanja regresionog i binarnog klasifikacionog modela. Mape doprinosa atoma su izgrađene primenom koncepta mapa sličnosti gde se doprinos svakog atoma računa na osnovu razlike u predviđenim verovatnoćama dobijenim sa, odnosno bez uklanjanja bitova fingerprintova koji se odnose na specifični atom. U cilju lakšeg dizajna novih inhibitora od strane korisnika, u *SIRT2i_Predictor* uključena je i analiza Tanimoto sličnosti (ECFP4) ispitivanih jedinjenja u odnosu na sve do danas poznate SIRT2 inhibitore. Ova informacija se takođe može upotrebiti i pri proceni pouzdanosti predviđanja modela selektivnosti, naročito DNN:deskriptorskog SIRT2/3 modela koji je pokazao lošiju statistiku na setu molekula mamaca strukturno raznorodnih jedinjenja. Ovaj modul kao i VS

modela, nedvosmisleno je ukazala na tiohidantoinjski skelet kao deo strukture od ključnog značaja za aktivnost izdvojenog jedinjenja. Nijedan od tiohidantoinjskih derivata do danas nije prijavljen kao sirtuinski inhibitor čime je demonstriran potencijal *SIRT2i_Predictor*-a da izdvoji hemijski nove strukture i postavi osnovu hipotezu odnosa strukture i aktivnosti. Značajna prednosti *SIRT2i_Predictor*-a je mogućnost analiziranja 200000 jedinjenja za svega nekoliko minuta u odnosu na SBVS pristup koji zahteva nekoliko sati.



Slika 4.20. Rezultati studije poređenja. (A) Poređenje hemijskog prostora koji je pokriven inhibitorima SIRT2 iz ChEMBL baze (levo), virtuelnim *hit* jedinjenjima SBVS-a (sredina) i virtuelnim *hit* jedinjenjima *SIRT2i_Predictor*-a (desno); (B) analiza verovatnoća dobijenih za virtuelna *hit* jedinjenja izdvojena iz jedinstvenog dela hemijskog prostora koji je pokriven *SIRT2i_Predictor*-om (znak zvezdice u Slici A); (C) odnosi struktura-aktivnost izračunati koristeći mape sličnosti za binarni model (gornji grafik) i regresioni model (donji grafik) za virtuelne rezultate izdvojene iz jedinstvenog dela hemijskog prostora (znak zvezdice u Slici A). Crvenom bojom označen je region koji ima pozitivan uticaj na aktivnost, dok plava boja označava negativan uticaj.

Daljim poređenjem *SIRT2i_Predictor*-a sa SBVS pristupom, analizirano je 9 jedinjenja eksperimentalno testiranih u studijama validacije SBVS protokola. IC_{50} vrednosti 2 izdvojena *hit* jedinjenja i Inh% 5 jedinjenja su pripadala u tzv. „sivoj zoni“ prema pravilima klasifikacije binarnog modela ($IC_{50} = 50 - 90 \mu M$; Inh% $@200 \mu M = 40\% - 80\%$), dok su 2 jedinjenja bila inaktivna. Bitno je podsetiti da je koncept „sive zone“ uveden u dizajn studije radi smanjenja uticaja varijabilnosti unutar seta podataka koja je posledica različitih eksperimentalnih uslova *in vitro* testiranja, zbog čega su neka od aktivnih jedinjenja isključena iz modelovanja. Analizom ovih 9

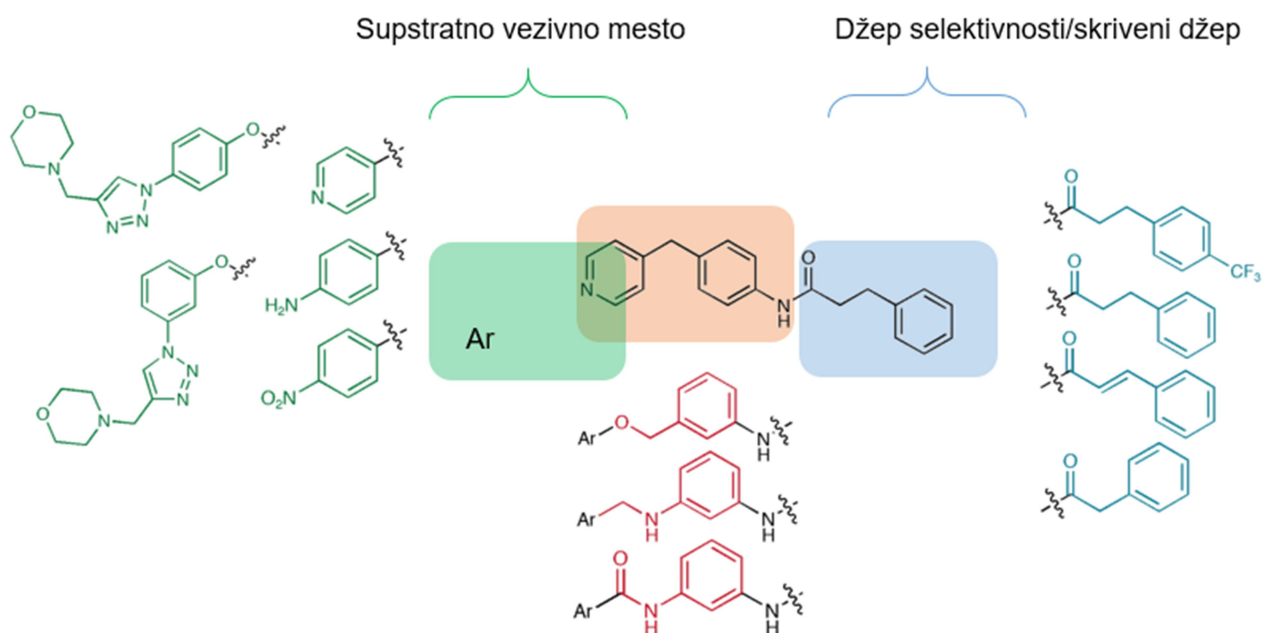
jedinjenja primenom *SIRT2i_Predictor*-a uočeno je slaganje sa eksperimentalnim rezultatima (Prilog 7.9). Naime, binarni klasifikacioni model predvideo je da nijedno od 9 jedinjenja ne pripada klasi aktivnih jedinjenja ($IC_{50} < 50 \mu M$) prema pravilima klasifikacije binarnih modela. Pet od devet jedinjenja je bilo izvan domena primenljivosti klasifikacionog modela. Prema načinu definisanja domena primenljivosti klasifikacionog modela, jedinjenja izvan ovog domena su jedinjenja sa graničnom klasifikacionom verovatnoćom. Interesantno je bilo uočiti da su sva jedinjenja sa graničnom verovatnoćom klasifikacije zapravo pripadnici „sive zone“ što je i očekivano s obzirom da je model treniran da predviđa aktivnosti van ove zone. Preostala 4 jedinjenja su predviđena kao inaktivna što je u skladu sa pravilima dodeljivanja klasa binarnog modela. Na ovaj način delimično je demonstrirana i sposobnost *SIRT2i_Predictor*-a za post analizu rezultata drugih VS pristupa (Prilog 7.9).

Dalja validacija *SIRT2i_Predictor*-a izvedena je analizom male interne baze jedinjenja. Nažalost, binarni model je predvideo da su sva jedinjenja bila van definisanog domena primenljivosti, ili su predviđena kao inaktivna. U cilju daljeg opravdavanja *SIRT2i_Predictor*-a kao validnog filtera u procesuiranju i analizi rezultata ostalih VS protokola, 4 jedinjenja predviđena kao aktivna primenom prethodno razvijenog SBVS pristupa testirana su *in vitro*. Iako su jedinjenja pokazala neki nivo SIRT2 aktivnosti, sva jedinjenja su se pokazala kao slabi inhibitori (3 jedinjenja su bila u „sivoj zoni“ dok je 1 jedinjenje bilo u kategoriji neaktivnih jedinjenja na osnovu pravila raspodele jedinjenja binarnog modela). Kao standardni inhibitor upotrebljen je EX-527 (IC_{50} (SIRT2) = 20 μM)³²⁵ koji nije bio uključen u trening set binarnih modela. *SIRT2i_Predictor* je uspeo da napravi jasnu distinkciju između EX-527 i testiranih jedinjenja čime je dodatno potvrđen potencijal primene ovog alata kao dodatnog filtera u studijama virtuelnog skrininga (Prilog 7.9).

U zaključku, *SIRT2i_Predictor* je u pogledu pokrivenosti hemijskog prostora virtuelnih *hit* jedinjenja pokazao uporedive rezultate u odnosu na računarski i vremenski zahtevnija SBVS protokol. Rezultati su ukazali na praktični značaj primene *SIRT2i_Predictor*-a kao samostalnog LBVS alata u otkriću novih selektivnih SIRT2 inhibitora, ali i kao dodatnog filtera za prioritizaciju jedinjenja u virtuelnim, ili eksperimentalnim skriningima velikih baza jedinjenja. Kao što je istaknuto iznad, *SIRT2i_Predictor* je razvijen uz grafički interfejs koji umnogome olakšava i omogućava primenu predstavnicima šire naučne javnosti.

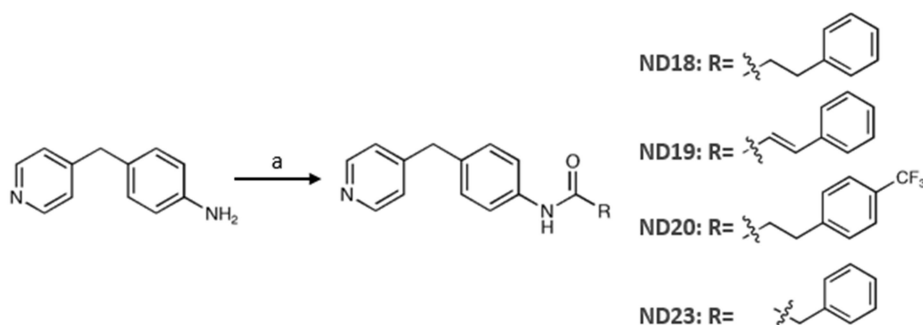
4.3. Sinteze novih SIRT2 inhibitora na osnovu strukture ND18

Primenom razvijenih računarskih modela i postavljanjem hipoteze o načinu interakcije *hit* jedinjenja ND18 sa SIRT2, dizajnirano je i sintetisano 12 novih SIRT2 inhibitora. Dizajn je započeo oslanjajući se na hipotezu o načinu interakcije ND18 generisanu od strane multistrukturalnog SBVS modela (Slika 4.21). U početnim fazama, dizajn je bio baziran isključivo na multistrukturalnom SBVS modelu pri čemu su jedinjenja ND18-20 i ND23 bila dizajnirana kao direktni analozi jedinjenja ND18 uz zadržavanje osnovnog fragmenta 4-(piridin-4-il-metil)anilina (Slika 4.22). Dizajn analoga ND18-20 i ND23 imao je za cilj ispitivanje uticaja različite dužine i rigidnosti fenil-alkil fragmenata na potentnost SIRT2 inhibicije 4-(piridin-4-il-metil)anilinskih derivata. Uzimajući u obzir da je SBVS model predvideo vezivanje fenil-alkil fragmenta ND18 u unutrašnjosti vezivnog mesta (tzv. džep selektivnosti, odnosno novootkriveni skriveni džep) uz samu ivicu unutrašnjeg zida vezivnog mesta, fragmenti duži od 3-(4-trifluormetil-fenil)-propionil grupe nisu dalje razmatrani u dizajnu. Preostala jedinjenja (ND24B – 24_D i ND3_7 – 7_7, Slike 4.23 – 4.25) su dizajnirana izosternim zamenama i proširenjem 4-(piridin-4-il-metil)anilinskog fragmenta ka spoljašnjosti vezivnog mesta, odnosno ka supstratnom vezivnom mestu, oslanjajući se i na predviđanja *SIRT2i_Predictor*-a. Kombinovanjem dve razvijene i prethodno validirane metodologije dizajna (multistrukturalni SBVS i *SIRT2i_Predictor*) bila je omogućena efikasnija pretraga inovativnih hemijskih vrsta SIRT2 inhibitora uz računanje sintetske izvodljivosti u datim uslovima rada. Dizajn derivata aril-1,2,3-triazola (jedinjenja ND3-7_7, Slike 4.24 i 4.25) je imao za cilj još jedan iskorak u neistraženi deo hemijskog prostora SIRT2 inhibitora primenom razvijenih računarskih modela.

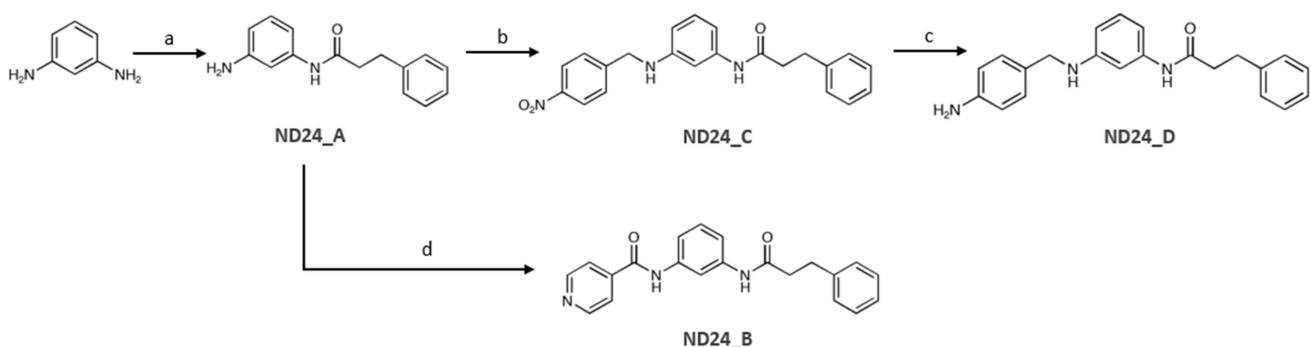


Slika 4.21. Osnovni princip dizajna analoga *hit* jedinjenja ND18 sa naznačenim predviđenim mestima vezivanja unutar SIRT2 vezivnog mesta inhibitora.

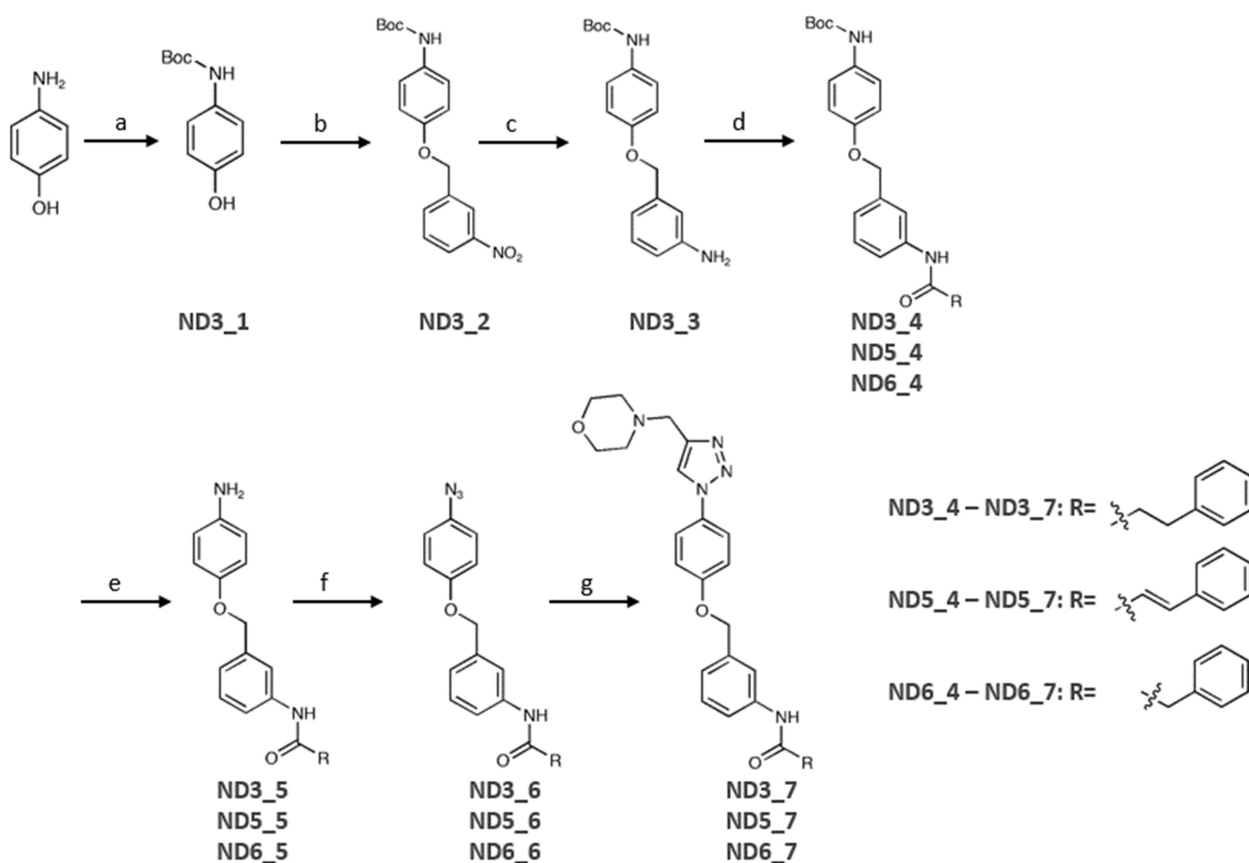
Analizom ChEMBL baze nije uočeno postojanje sličnih derivata među SIRT2 inhibitorima (Tanimoto koeficijent sličnosti <0,35 računato za 4-({1-[3-(benziloksi)fenil]-1,2,3-triazol-4-il}metil)morfolin). Sinteze dizajniranih inhibitora su izvedene prema optimizovanim procedurama predstavljenim na Slikama 4.22 – 4.25, uz prinose pojedinačnih reakcija navedenih u opisima slika. Sintetisana jedinjenja su okarakterisana NMR spektroskopijom (^1H NMR i ^{13}C NMR eksperimentima). U okviru Priloga 7.10 sumirani su spektri intermedijera i finalnih jedinjenja (hemijska pomeranja su prikazana u δ jedinicama (ppm), a konstante kuplovanja (J) u hercima (Hz), uz naznačene deuterio rastvarače upotrebljene za snimanja).



Slika 4.22. Procedura sinteze jedinjenja ND18-20 i ND23. Reakcija a – DCM/DMF (10:0,5), HOBt, EDCI, TEA uz dodatak fenil-propanske (ND18), cimetine (ND19), 4-trifluormetil-fenil-propanske (ND20), ili fenil-sirćetne (ND23) kiseline, sobna temperatura, 24h (prinos 32,5-37,8%).

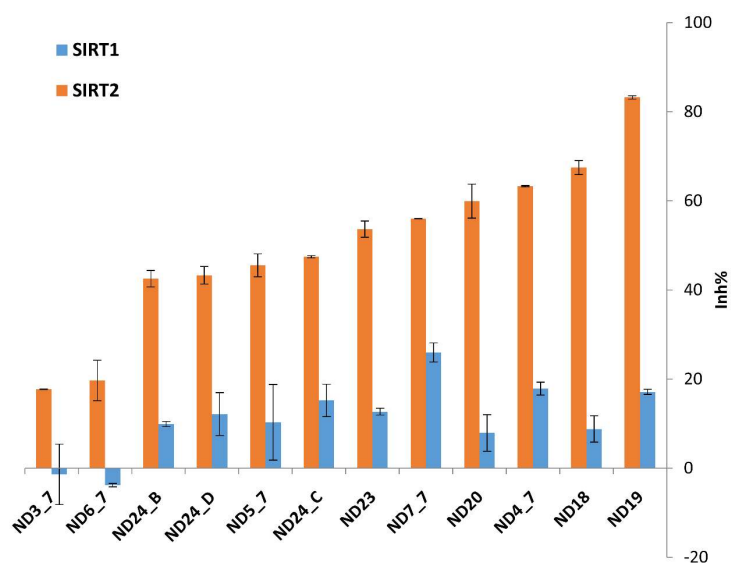


Slika 4.23. Procedura za sintezu jedinjenja ND24_B, ND24_C i ND24_D. Reakcije a – DCM/DMF (10:0,5), HOBT, EDCI, TEA uz dodatak fenil-propanske kiseline (prinos 37,8%); Reakcija b – DCM, 4-nitrobenzalhid, *Hantzsch*-ov estar, TFA, anhidrovani $MgSO_4$ i molekulska sita, refluks 12h (prinos 64%); Reakcija c – ACN/ H_2O (3:0,5), $NiCl_2 \cdot 6H_2O$, $NaBH_4$, sobna temperatura, 1h (prinos 89%); Reakcija d – DCM/DMF (10:0,5), HOBT, EDCI, TEA uz dodatak izonikotinske kiseline (prinos 40,5%).



Slika 4.24. Procedura za sintezu jedinjenja ND3_4, ND5_7 i ND6_7. Reakcija a – THF, di-terc-butil dikarbonat, sobna temperatura, 24h (prinos > 90%); Reakcija b – ACN, K_2CO_3 , 3-nitrobenzilbromid, sobna temperatura, 24h (prinos > 90%); Reakcija c – ACN/ H_2O (3:0,5), $NiCl_2 \cdot 6H_2O$, $NaBH_4$, sobna temperatura, 1h (prinos > 90%); Reakcija d – ACN, HOBT, EDCI, DMAP, DIPEA uz dodatak fenil-propanske (ND3_4), cimetine (ND5_7), ili fenil-sirćetne (ND6_7) kiseline, sobna temperatura, 24h (prinosi redom 48,5%, 42,3% i 56,7%); Reakcija e – DCM:TFA (25%), sobna temperatura (prinos > 90%); Reakcija f – $CH_3COOH:H_2O$ (5:2), $NaNO_2$ (1,3 M rastvor u H_2O),

potencijala. Uprkos visoko rizičnom pristupu u dizajnu derivata aril-1,2,3-triazola, dobijene inhibitorne aktivnosti ND4_7 i ND7_7 ohrabruju nastavak rada na optimizaciji strukture i aktivnosti ove potpuno nove grupe inhibitora.



Slika 4.26. Procenti inhibicije aktivnosti SIRT1 i SIRT2 pri koncentraciji od 100 μ M testiranih jedinjenja.

4.4. Molekulsko modelovanje protein-ligand interakcije fotoreaktivnih peptidomimetskih inhibitora WDR5

Istraživanja predstavljena u ovom delu disertacije su izvršena u cilju postavljanja protokola za ispitivanje mehanizama protein-ligand interakcije kao i odnosa strukture i aktivnosti četiri predstavnika potpuno nove grupe fotoreaktivnih peptidomimetskih inhibitora WDR5, od kojih su tri iz grupe linernih i jedan iz grupe cikličnih peptidomimetika (Slika 3.4 i poglavlje 1.4.2.a). Svi inhibitori su izučavani u dva stanja – *cis* i *trans* što čini ukupno osam izučavanih sistema.

4.4.1. Molekulsko modelovanje protein-ligand interakcije linearnih peptidomimetskih inhibitora WDR5

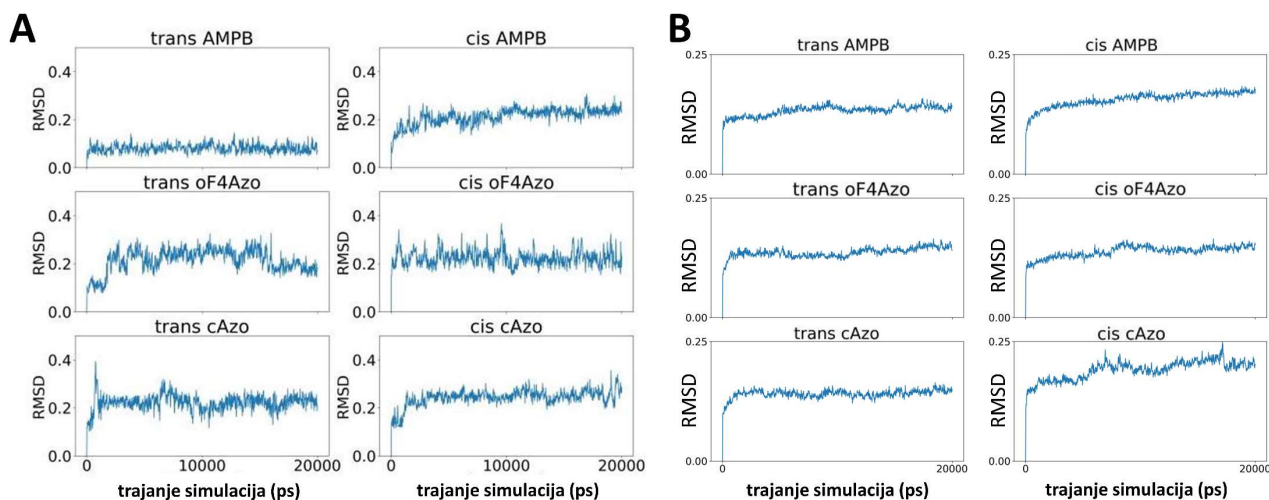
Selektovani inhibitori predstavljaju analoge peptidne sekvence MLL1 koja sadrži WIN motiv (tzv. WIN peptid Ac-GSARAEVHLRKS-NH₂) sa inkorporiranom fotoreaktivnom azobenzenskom grupom³²⁶. Kako bi se stekao atomistički uvid u prirodu protein-ligand interakcije, i postavio protokol za dizajn novih fotoreaktivnih linearnih peptidomimetika, izvršene su serije eksperimenata molekulskog dokinga i MD simulacija. Za svih šest liganda (Slika 3.4): *cis/trans* AMPB-peptid (3), *cis/trans* cAzo-peptid (4), *cis/trans* oF4Azo-peptid (5), početne konfiguracije sistema za MD simulacije dobijene su molekulskim dokingom.

Klasični/tradicionalni doking programi su prvenstveno razvijani za doking malih molekula i benčmarking studije ukazuju da njihova pouzdanost drastično opada pri dokingu velikih fleksibilnih molekula poput peptidomimetika²⁹¹. Jedan od osnovnih problema sa molekulskim dokingom peptida i peptidomimetika predstavlja ogroman broj stepeni slobode sa čijim porastom se eksponencijalno usložnjava i problem koji algoritmi za pretragu vezivnih konformacija moraju da reše²⁹⁰. Jedna od strategija kojom se smanjuje kompleksnost problema koji rešavaju algoritmi pretrage predstavlja fragmentacija molekula peptida/peptidomimetika. Fragmentisanjem strukture peptida/peptidomimetika problem dokinga se svodi na rešavanje poze dva, ili više uslovno malih molekula.³²⁷ Uspešnost ove strategije umnogome zavisi i od predznanja o načinu interakcije peptida sa vezivnim mestom pri čemu implementacija ovih informacija u doking protokol može značajno suziti pretragu i povećati efikasnost dokinga. U slučaju dokinga serije linearnih peptidomimetika, informacije o načinu interakcije -**ARA**- (-alanin-arginin-alanin-) segmenta sa WDR5 dostupne iz kristalografskih studija poslužile su za postavljanje ograničenja za pretragu bioaktivnih konformacija liganada.

Treba napomenuti da se, usled povećanja popularnosti peptidnih i peptidomimetskih lekova, poslednjih godina razvijaju i specijalizovani programi za doking većih peptida. Međutim, većina razvijenih programa nije namenjena peptidima/peptidomimeticima sa nestandardnim reziduama (npr. azobenzenske rezidue) i njihove performanse su samo umereno povećane u odnosu na klasične doking programe, usled čega je doking protokol istraživanja ovog dela doktorske disertacije postavljen primenom GOLD programa (poglavlje 3.2.1). Doking protokol je validiran redokingom *trans* AMPB-peptida (3) u kristalografski razrešenoj strukturi (PDB kod: 5M23) i računanjem korena srednjeg kvadratnog odstupanja (RMSD) koordinata teških atoma za deo *trans* AMPB-peptida. Dobijena vrednost RMSD od 0,8915 ukazala je na validnost prikazanog protokola.

Uzimajući u obzir ograničenja postojećih metoda za molekulski doking peptidomimetika sa nestandardnim reziduama, razvijeni protokol za molekulski doking je bio udružen sa simulacijama molekulske dinamike u cilju provere i korigovanja struktura kompleksa dobijenih molekulskim dokingom. Imajući u vidu značajnu konformacionu fleksibilnost peptidomimetika, generisano je nekoliko početnih kompleksa za svaki ligand. Samo kompleksi proteina i liganda u kojima se *N*-terminalni deo peptidomimetika (SARA-) održao stabilnim tokom 20 ns MD simulacije korišćeni su za dalju analizu. Rezultati MD simulacija izraženi kao koren srednjeg kvadratnog odstupanja (RMSD) fluktuacija *N*-terminalnog dela peptidomimetika (SARA-) i peptidnog skeleta WDR5

(Slika 4.27) tokom 20 ns produkcione MD simulacije ukazuju na stabilizovane komplekse koji su konvergirali u stanje ravnoteže.



Slika 4.27. Rezultati RMSD fluktuacija SARA sekvence izučavanih liganada tokom 20 ns simulacija.

4.4.1.a. MM/PBSA proračuni

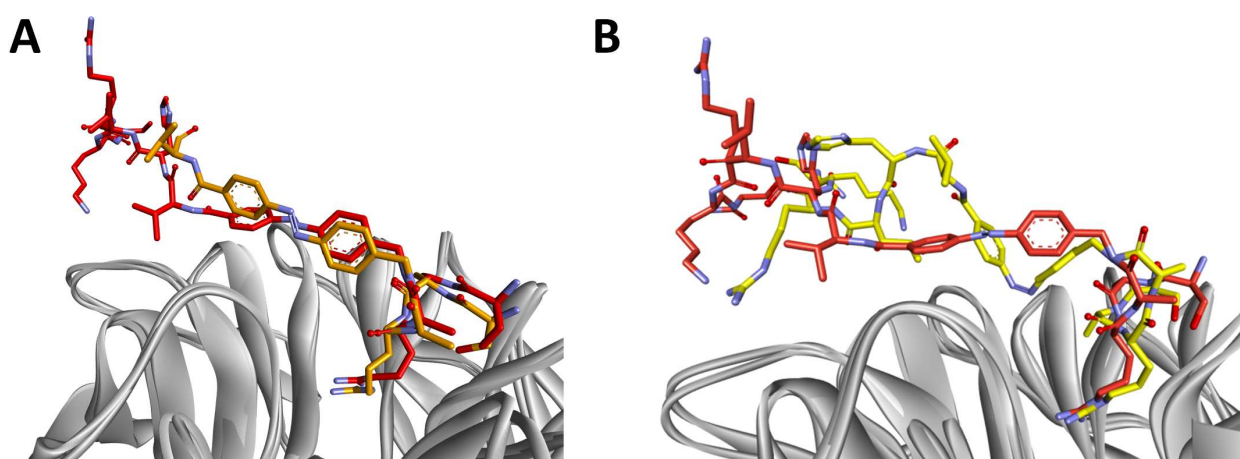
Proračuni slobodne energije vezivanja primenom MM/PBSA metode (eng. *Molecular mechanics Poisson–Boltzmann surface area* – MM/PBSA) izvedeni su u okviru poslednjih 7 ns svake od izvedenih simulacija u cilju dalje validacije predviđenih načina interakcije peptidomimetika. Upotrebom rigoroznije scoring metode kao što je MM/PBSA bilo je moguće korelisati eksperimentalne K_i vrednosti²¹³ (Tabela 4.7) (transformisane pomoću relacije $pK_{i(\text{exp})} = -\log K_i$) sa predviđenim slobodnim energijama vezivanja i na taj način dodatno validirati predviđene vezivne poze. Ovakav vid proračuna nije bilo moguće izvesti sa skorovima dobijenim iz molekuskog dokinga zbog primene metode fragmentacije. Linearni regresioni model sa $R^2 = 0,88$ između navedenih parametara (Tabela 4.7), ukazao je na visok stepen korelacije eksperimentalnih aktivnosti sa predviđenim MM/PBSA skorom. Literaturni podaci ukazuju da MM/PBSA, iako namenjen proračunima slobodne energije vezivanja, zbog visokog nivoa uvedenih aproksimacija ima širu primenu kao scoring funkcija, nego kao alat za analizu apsolutnih i relativnih slobodnih energija vezivanja⁷⁴. U konkretnom slučaju, nije bilo moguće uspostaviti direktnu vezu između K_i vrednosti i MM/PBSA slobodne energije vezivanja čak ni u smislu relativnih odnosa K_i i izračunatih energija. Razlog za ovo bi se mogao pronaći u velikoj konformacionoj fleksibilnosti peptidomimetika (diskutovano ispod) i nemogućnošću kratkih MD simulacija da posete sve alternativne vezivne modove liganda. Sa druge strane MM/PBSA se kao skor pokazao odličan u diskriminaciji peptidomimetika koji se vezuju sa većim ili manjim afinitetom što je najverovatnije u vezi sa visokim udelom ispitivanog konformacionog stanja u ukupnoj energiji vezivanja. S obzirom da se razvijeni protokol oslanja na relativno nizak utrošak računarskih resursa (molekulski doking i kratke MD simulacije) pri čemu uspeva da rangira derivate peptidomimetskih inhibitora po aktivnosti što opravdava dalju upotrebu razvijenog protokola u dizajnu novih predstavnika.

Tabela 4.7. Rezultati MM/PBSA reskoringa.

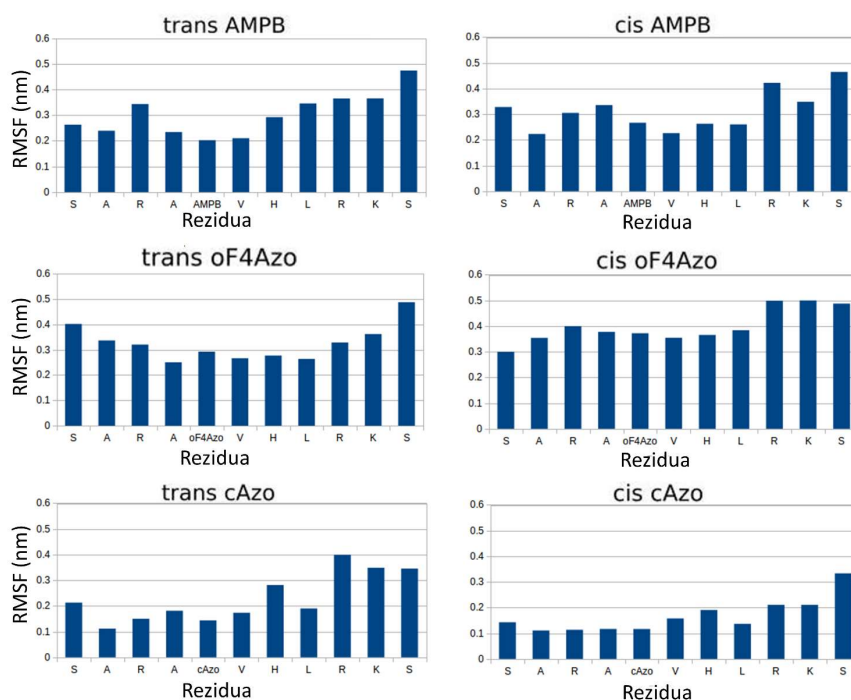
Peptidomimetik	pK _i (_{exp})	MM/PBSA skor (kJ/mol)
<i>trans</i> AMPB-peptid (3)	8.903	-51.282
<i>cis</i> AMPB-peptid (3)	8.187	-21.136
<i>trans</i> oF4Azo-peptid (5)	7.923	-28.388
<i>cis</i> oF4Azo-peptid (5)	7.511	-15.371
<i>trans</i> cAzo-peptid (4)	6.854	-5.909
<i>cis</i> cAzo-peptid (4)	6.684	-5.680
$pK_{i(\text{exp})} = -19.2458(\text{MM/PBSA}) + 126.4877$		
$R^2 = 0.88$		

Kako bi se stekao uvid u detalje intermolekulskih interakcija WDR5 i liganda, nakon postupka klasterovanja trajektorija, dobijeni centroidi klastera su dalje analizirani. Za sve ligande, najveće fluktuacije uočene su u C-terminalnom delu peptidomimetika (-VHLRKS) (Slika 4.29). Ovi rezultati su bili u skladu sa prethodno objavljenom kristalografskom studijom kompleksa AMPB-peptid-WDR5 (PDB kod: 5M23), u kojoj elektronske gustine C-terminalni deo peptidomimetika (-VHLRKS) nije bilo moguće razrešiti²¹².

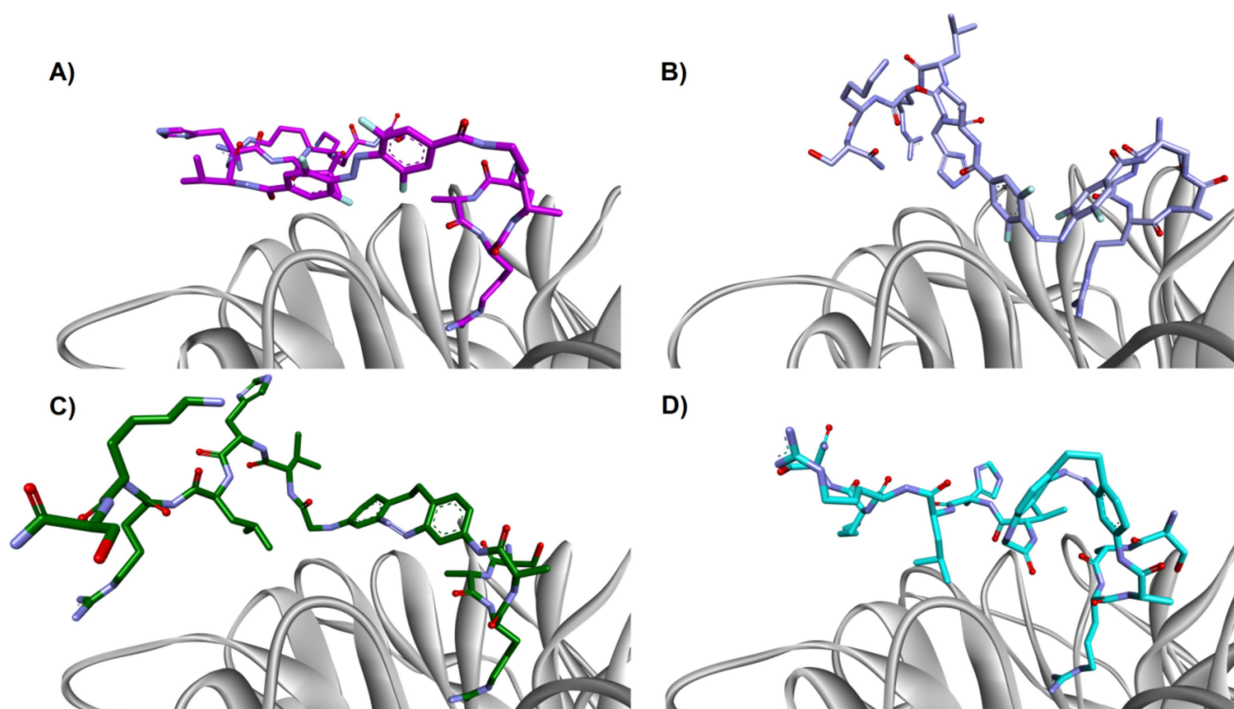
Način vezivanja *N*-terminalnog dela *trans* AMPB-peptida (3) tokom celog toka MD simulacija ostao je sličan načinu vezivanja opisanom u kristalnoj strukturi (PDB kod: 5M23) (Slika 4.28), što dodatno potvrđuje postavljenu računarsku metodologiju. Bolja stabilizacija SARA-sekvenca *trans* AMPB-peptida (3) u poređenju sa ostalim ligandima potvrđena je RMSF dijagrama RMSF (Slika 4.29). Prema MD simulacijama *trans* AMPB-peptida (3), valin, kao najmanje fluktuirajući ostatak iz C-terminalnog dela peptidomimetika (-VHLRKS) (Slika 4.29), prepoznat je kao važan ostatak za stabilizaciju proteina-ligand kompleksa. Upoređujući položaje *cis* AMPB-peptida (3) i *trans* AMPB-peptida (3) (Slika 4.28), *N*-terminalni deo oba peptidomimetika (SARA-) zauzeo je sličnu konformaciju, dok je najveća razlika bila u orijentaciji rezidua u C-terminalnom delu (-VHLRKS) (Slika 4.28).



Slika 4.28. Molekulski modeli AMPB-peptid (3):WDR5 kompleksa. A – Poređenje strukture WDR5 kompleksa sa *cis* AMPB-peptidom (3) dobijenog kroz MD simulacije (crveno) i razrešenog dela liganda iz kristalografskih studija (narandžasto) PDB ID: 5M23; B – Razlika u načinu interakcije sa WDR5 predviđenim za *trans* AMPB-peptid (3) (crvena) i *cis* AMPB-peptid (3) (žuta) unutar veznog mesta WDR5.



Slika 4.29. RMSF fluktuacije rezidua peptidomimetskih inhibitora tokom MD simulacija.



Slika 4.30. Poravnjane strukture modelovanih kompleksa linearnih peptidomimetika sa WDR5. A – *trans* oF4Azo-peptid (5) (tamnoljubičasta); B – *cis* oF4Azo-peptid (5) (svetloljubičasta); C – *trans* cAzo-peptid (4) (zelena); D – *cis* c-Azo peptid (tirkizna).

Način vezivanja *trans* oF4Azo-peptida (5) ostao je očekivano sličan načinu vezivanja *trans* AMPB-peptida (3) u pogledu *N*-terminalnog domena, dok je tetra-orto-fluoroazobenzenska grupa delimično pomerenjena u odnosu na azobenzensku grupu *trans* AMPB-peptida (3) (Prilog 7.11). Leucin u C-terminalnom delu (-VHLRKS) je pokazao prolazne intramolekulske interakcije sa tetra-orto-fluoroazobenzenskom grupom tokom MD simulacije (Prilog 7.11). Ova specifična intramolekulska

interakcija je bila glavni razlog za povećane fluktuacije C-terminalnog valina u odnosu na isti valin u *trans* AMPB-peptidu (3) tokom MD simulacija (Slika 4.29). Valin je zadržao sličan način vezivanja unutar WDR5 vezujućeg mesta u poređenju sa *trans* AMPB-peptidom (3) (Prilog 7.11). Međutim, slična orijentacija i pozicija C-terminalnog valina kod *trans* peptida 3 i 5 ukazuju na njegovu važnost u ukupnoj stabilizaciji kompleksa proteina i liganda. Primećen je značajan pomak u vezivnoj konformaciji SARA-domena u slučaju *cis* oF4Azo-peptida (3) (Slika 4.30 i Prilog 7.11). Razlika u posmatranim načinima vezivanja između *cis/trans* AMPB-peptida (3) i *cis/trans* oF4Azo-peptida (5) može se pripisati različitoj dužini linkera između SARA-domena i halogenovanog azobenzena (5). Naime, peptidomimetički molekul (5) kao linker poseduje 2-aminoetil-karbamoil koji povezuje halogenovani azobenzen i SARA-domen, dok kod peptidomimetika (3) metilenska grupa je ta koja povezuje SARA-domen i nesupstituisani azobenzen.

Proračuni molekulske dinamike izvedeni za *trans* cAzo-peptid (4) ukazuju da ovaj izomer samo delimično oponaša pozicioniranje drugih proučavanih *trans*-liganda (Slika 4.30C i Prilog 7.11). Kristalografska struktura kompleksa WDR5 sa *cis* cAzo-peptidom (4) je razrešena tek nakon izvedene studije molekuskog modelovanja. U pogledu interakcije SARA sekvence i cikličnog azobenzena sa WDR5, predviđeni način vezivanja *cis* cAzo-peptida (4) je bio uporediv sa naknadno razrešenom kristalnom strukturom (PDB kod: 6IAM) čime je dodatno potkrepljena validnost razvijenog računarskog protokola (Prilog 7.11). Izrazita stabilnost SARA sekvence tokom simulacije ne iznenađuje s obzirom da je ovaj deo strukture u studijama odnosa strukture WIN peptida i aktivnosti definisan kao uslov za dejstvo. Međutim, uočene su određene razlike u orijentaciji C-terminalne sekvence –VHLRKS koja je tokom simulacija u svim izučavanim sistemima, pa i u ovom, pokazala fleksibilnost u načinima interakcije sa površinom WDR5. Bez obzira na uočena odstupanja od kristalne strukture, interakcija sa aromatičnim reziduama Tyr-131, Phe-149 ostala je stabilna tokom celog toka simulacije (Prilog 7.11). Studije mutageneze su ukazale na ove dve rezidue kao izrazino značajne u stabilizaciji interakcije sa peptidima.³²⁸ U poređenju sa ostalim simuliranim kompleksima, C-terminus *cis* cAzo-peptida (4) je interagovao sa manjim delom površine WDR5. Ovakvo ponašanje se može pripisati kompaktnijoj strukturi celog liganda. Studije molekulske dinamike su ukazale na visoku fleksibilnost C-terminalnih sekvenci ostalih liganada što može pružiti i objašnjenje za nemogućnost razrešavanja strukture kompleksa primenom kristalografske analize koja je prijavljena tokom izvođenja studije molekuskog modelovanja.

4.4.2. Molekulsko modelovanje protein-ligand interakcije cikličnog peptidomimetskog inhibitora WDR5

U nastavku istraživanja izvedenih u okviru ove disertacije, izučavan je mehanizam dejstva najpotentnijeg predstavnika (CS-VIP-8) potpuno nove grupe cikličnih fotoreaktivnih inhibitora WDR5 (Slika 3.4). Osnovna prednost CS-VIP-8 u odnosu na prethodno opisane linearne peptide predstavlja visok odnos selektivnosti dva fotostacionarna stanja *cis/trans*, pri čemu za razliku od linearnih peptidomimetika *cis* forma aktivnija²¹⁴.

Osnovne prednosti cikličnih peptidomimetskih inhibitora u odnosu na linearne peptidomimetike mogu se predstaviti kroz sledeće tačke³²⁹:

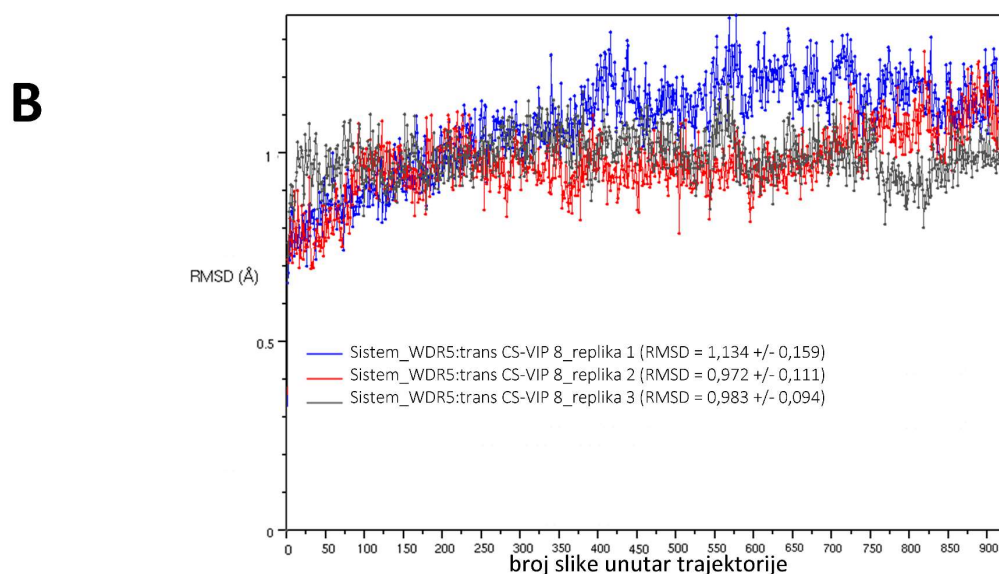
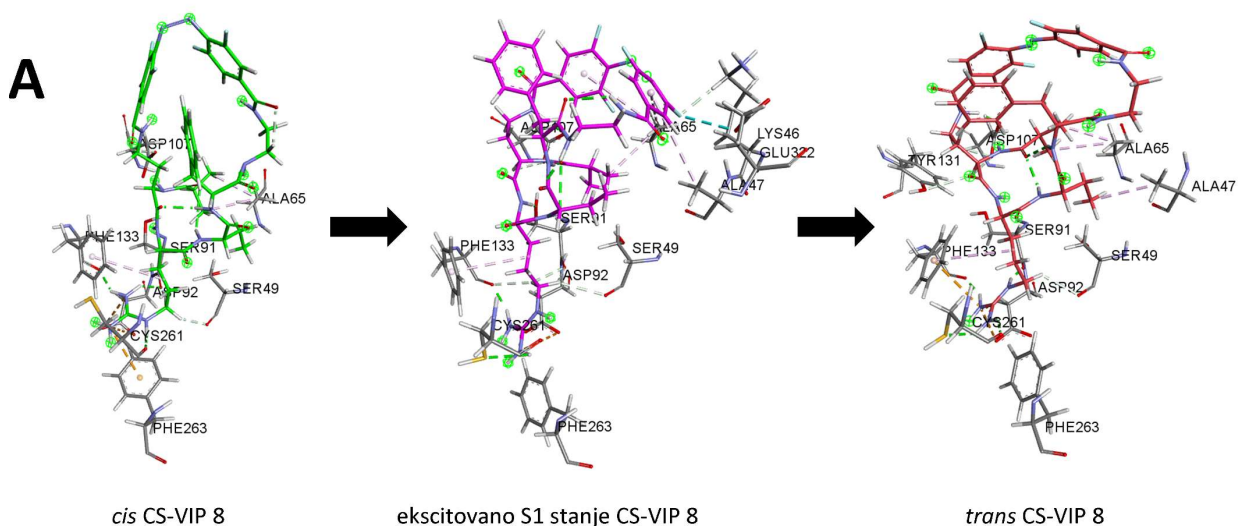
- Povećana stabilnost – Zatvorena struktura cikličnih peptida povećava njihovu stabilnost u odnosu na linearnu strukturu. Ova stabilnost može pomoći u zaštiti peptida od razgradnje i raspadanja;
- Veća specifičnost – Ciklizacijom i konformacionom rigidifikacijom se postiže veća specifičnost ka ciljnom mestu dejstva cikličnih peptidomimetika, što za posledicu ima i manji broj neželjenih efekata;
- Bolje farmakokinetičke osobine i povećana bioraspoloživost – Ciklični peptidi poseduju manju osetljivost na proteolitičke enzime zbog čega se očekuje i veća bioraspoloživost nakon oralne primene. Ciklični peptidi mogu biti dizajnirani da imaju bolje

farmakokinetičke osobine u odnosu na linearne peptidne ligande. Ovo može uključivati duže poluvreme eliminacije i veću oralnu apsorpciju.

4.4.2.a. Molekulsko modelovanje WDR5:*trans*-CS-VIP 8 kompleksa i studija fotoizomerizacije

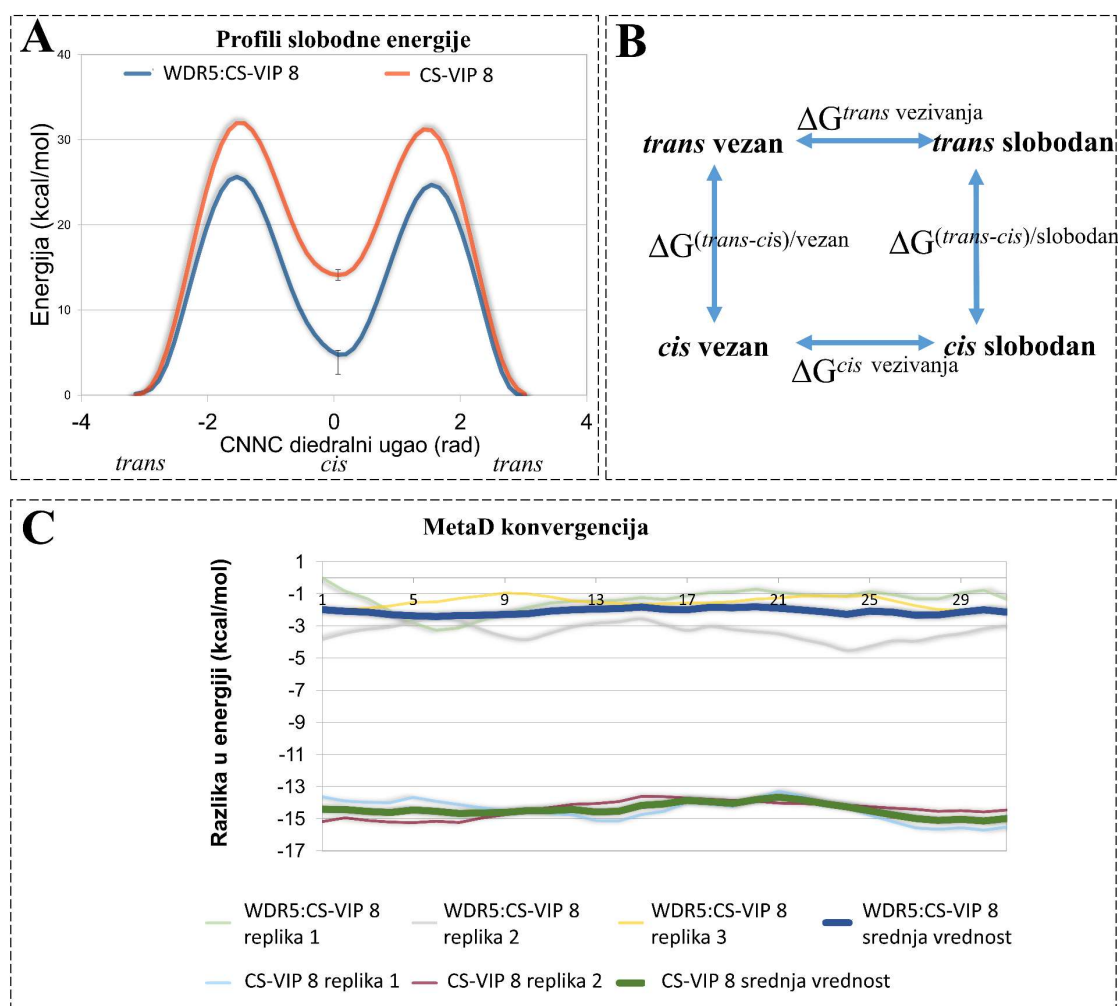
Eksperimentalnim metodama dobijeni su samo ograničeni i podaci o mehanizmu fotoizomerizacije i interakciji *trans* izomera CS-VIP 8 inhibitora sa WIN vezivnim mestom na WDR5. Uprkos činjenici da su eseji bazirani na polarizacionoj fluorescenciji ukazali da se *cis* izomer vezuje skoro 10 puta većim afinitetom za WDR5 (odnos selektivnosti *cis:trans* stanja inhibitora od 1:9,88), zbog ograničenja eksperimentalnih merenja odnosno mogućnosti nepotpune fotoizomerizacije, nije bilo moguće sa sigurnošću utvrditi koliki je udeo rezidualnog *cis* stanja u merenjima izvršenim za fotostacionarno *trans* stanje²¹⁴. Dalje, merenja afiniteta tokom *in situ* izomerizacije ukazala su na mogućnost otežanog procesa prelaska *cis* u *trans* stanje u prisustvu WDR5. Treba istaći da kristalografskom studijom, za razliku od *cis* stanja, strukturu WDR5:*trans*-CS-VIP 8 kompleksa nije bilo moguće razrešiti kristalizacijom u pristustvu fotostacionarnog *trans* liganda (zbog prisutva rezidualnog *cis* stanja većeg afiniteta), niti kristalizacijom uz *in situ* izomerizaciju *cis* u *trans* gde je gotovo uvek dolazilo do difuzije *trans* liganda iz aktivnog mesta²¹⁴. Struktura WDR5:*cis*-CS-VIP 8 kompleksa je uspešno razrešena što je poslužilo kao polazna tačka za detaljna istraživanja protein-ligand interakcije²¹⁴. U okviru istraživanja predstavljenih u ovom delu disertacije izučavane su promene slobodne energije vezivanja i konformacione promene povezane sa fotoizomerizacijom CS-VIP 8 unutar WIN vezivnog mesta WDR5. Korišćene su dve nezavisne metode: a) simulacija molekulske dinamike sa „dijabatskim promenljivim potencijalom“ i b) simulacija metadinamike dobro temperiranih sistema (MetaD).

Pristupu dijabatskog promenljivog potencijala je primenjen pre svega u cilju razrešavanja strukture WDR5:*trans*-CS-VIP 8 kompleksa. U okviru ovog pristupa, modifikacijom parametara molekulske mehaničkih polja sile kreirane su funkcije dijabatskog potencijala prebacivanja iz osnovnog S0 *cis* stanja u pobuđeno stanje S1 i prebacivanja iz S1 stanja sistema u osnovno S0 *trans* stanje. Nakon iniciranja pobuđenog S1 stanja, u kompleksu sa WDR5, sistem je simuliran do 25 ns (Slika 4.31). Difuzija liganda koja je primećena u kristalografskoj studiji tokom *in situ* izomerizacije nije primećena u simulacijama S1 stanja, verovatno zbog razlika u eksperimentalnim i simuliranim uslovima, kao i zbog ograničenog uzorkovanja MD simulacija. Dalja analiza pokazala je da je -D-Dab-Arg-Abu- region CS-VIP 8 ostao u bioaktivnoj konformaciji tokom simulacija, bilo u S1 stanju ili tokom primene prebacujućih potencijala, sa očuvanjem svih ključnih intra- i intermolekularnih veza (slika 4.31). Nakon primene prebacujućeg potencijala, sistem je vraćen nazad u S0 stanje koje ovog puta odgovara *trans*-CS-VIP 8 konfiguraciji. Od tog trenutka, sistem je simuliran standardnim, ravnotežnim simulacijama (100 ns MD simulacija bez prebacivanja potencijala, u tri ponavljanja), kako bi se procenila pouzdanost predviđene strukture kompleksa WDR5:*trans*-CS-VIP 8. Analiza RMSD-a (slika 4.31B) ukazala je na odgovarajuću stabilizaciju sistema. Međutim, devijacije unutar izračunatih RMSD bile su najveća među svim proučavanim sistemima, što bi moglo ukazivati na značajan pad afiniteta *trans* stanja CS-VIP 8 u poređenju sa *cis* stanjem. U cilju kvantifikacije afiniteta *cis* i *trans* stanja prema WDR5, primenjena je MetaD metoda.



Slika 4.31. Rezultati simulacije dijabatskog promjenjivog potencijala. A – Pregled celokupnog protokola i poređenje režima interakcije *cis*-CS-VIP 8 ko-kristalne strukture (levo – zelena boja), modelovanog S1 stanja WDR5:CS-VIP 8 (sredina - ljubičasta) i modelovanog WDR5:*trans*-CS-VIP 8 (desno - crvena); B – Fluktuacije RMSD vrednosti detektovane u tri ponavljanja MD simulacija *trans*-CS-VIP 8 sistema.

Kako bi se potvrdio način vezivanja *trans*-CS-VIP 8 i steklo dublje razumevanje energetike iza *cis/trans* izomerizacije, upotrebljena je MetaD metoda. Sposobnost ove metode da istraži promene slobodne energije povezane sa krajnjim stanjima termodinamičkog ciklusa korišćena je za opis uticaja koji WDR5 ima na proces *cis/trans* izomerizacije. Poređenje profila slobodne energije dobijenih pomoću MetaD za sam ligand CS-VIP 8 i kompleks WDR5:CS-VIP 8 otkrilo je značajne razlike u $\Delta G(\textit{trans-cis})$ između vezanog i slobodnog liganda (Slika 4.32A) što je u skladu sa zapažanjima iz eksperimenata *in situ* izomerizacije. Termodinamički ciklus korišćen u ovim proračunima prikazan je na Slici 4.32B.



Slika 4.32. Rezultati MetaD simulacija. A – Poređenje dobijenih profila slobodne energije *cis/trans* izomerizacije kompleksa WDR5:CS-VIP 8 (plava linija) i slobodnog CS-VIP 8 (narandžasta linija) prikazane u odnosu na torzionu ugao azobenzena CNNC. Linije sa greškom označavaju standardnu devijaciju merenja razlike slobodne energije između *cis* i *trans* stanja na svim replikama; B – Termodinamički ciklus korišćen za izračunavanje $\Delta\Delta G^{(trans-cis)}$ vezivanja ($\Delta\Delta G^{(trans-cis)}$ vezivanja = $\Delta G^{(trans-cis)}$ slobodan - $\Delta G^{(trans-cis)}$ vezan = -14.09 - (-4.75) = -9.35 kcal/mol); C – Konvergencija razlike slobodne energije između *cis* i *trans* stanja za kompleks WDR5:CS-VIP 8 (plava linija) i slobodni CS-VIP 8 (narandžasta linija) u poslednjih 40 ns MetaD simulacija.

Sistemi su simulirani sve dok nije postignuta konvergencija svih replika (Slika 4.32C), a dobijene slobodne energije *cis/trans* izomerizacije ($\Delta G^{(trans-cis)}$) otkrile su da je vezano stanje povezano sa drastično manjom promenom slobodne energije izomerizacije ($\Delta G^{(trans-cis)}$) u poređenju sa slobodnim ligandom. Ovaj rezultat je ukazao na to da WDR5 u većoj meri stabilizuje *cis* stanje u poređenju sa odgovarajućim *trans* stanjem. Kao posledica ovoga, *cis* stanje CS-VIP 8 ima drastično veći afinitet vezivanja za WDR5. Predviđena $\Delta\Delta G^{(trans-cis)}$ vezivanja = -9,35 kcal/mol ukazuje na nekoliko stotina hiljada puta veći afinitet *cis*-CS-VIP 8 stanja u uslovima simulacija (Slika 4.32A). Ovaj rezultat potkrepljuje prethodno diskutovano eksperimentalno zapažanje da izmereni afinitet vezivanja *trans*-CS-VIP 8 može poticati od preostalog *cis*-izomera²¹⁴.

Dodatno, prema dobijenim profilima slobodne energije, prisustvo WDR5 značajno povećava energetska barijeru za prelaz iz *cis* u *trans* stanje (Slika 4.32A). Ovaj rezultat je bio u skladu sa hipotezom generisanom na osnovu eksperimenata *in situ* izomerizacije kojom se pretpostavlja da prisustvo WDR5 značajno otežava *cis-trans* izomerizaciju. *Trans* izomer je bio termodinamički stabilniji u oba stanja, što je u skladu sa standardnim ponašanjem azobenzena²¹⁴.

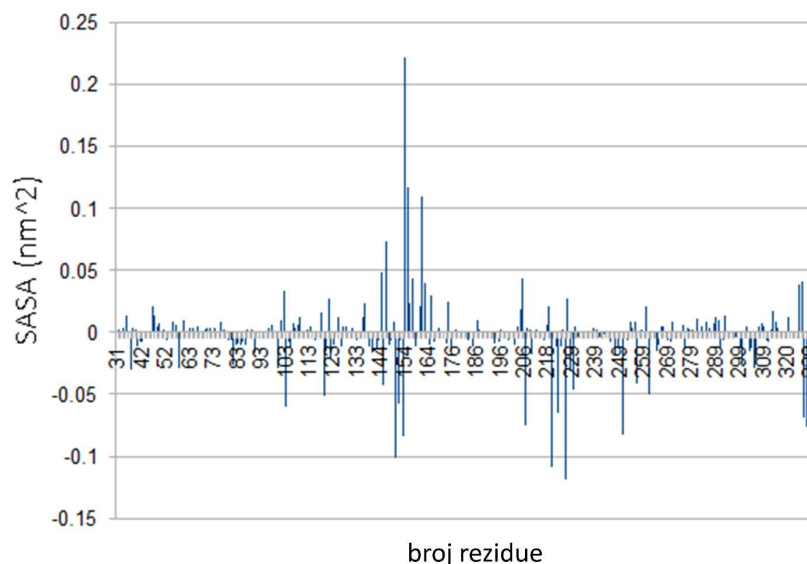
U pogledu načina interakcije *trans*-CS-VIP 8, predviđanja MetaD simulacije i dijabatskog prebacivanja su bili usaglašeni. Nakon klasterovanja svih dobijenih trajektora, centroid koji odgovara *trans*-CS-VIP 8 je upoređen sa ko-kristalnom strukturom *cis*-CS-VIP 8 stanja (Slika 4.31). Studija molekuskog modelovanja ukazala je na veliko konformaciono pomeranje liganda tokom fotoizomerizacije, ali je uprkos tome -D-Dab-Arg-Abu- region CS-VIP 8 zadržao konformaciju koja je ličila na bioaktivnu. Najveća razlika između *cis* i *trans* stanja je bila u pozicioniranju azobenzenskog dela CS-VIP 8. Međutim, ovo konformacijsko pomeranje je rezultiralo velikim padom afiniteta *trans*-izomera prema WDR5. Uzeti zajedno, rezultati MD simulacija dijabatskog prebacivanja i MetaD simulacija potvrđuju hipotezu o drastično većem afinitetu *cis* izomera i mogućnosti uticaja rezidualnog *cis* izomera na izmereni afinitet fotoreaktivnog *trans* stanja. Takođe, simulacije potvrđuju hipotezu o ulozi WDR5 u otežavanju procesa *cis-trans* fotoizomerizacije.

4.4.2.b. Analiza simulacija klasične MD

Mehanističke studije primenom eksperimenata vodonik-deuterijum izmene (eng. *Hydrogen-deuterium exchange* – HDX), izvedene na WDR5 i multiproteinskom MLL1 kompleksu u prisustvu *cis*-CS-VIP 8, ukazale su na određene delove u strukturama proteina koji su naročito pogođeni prisustvom liganda kroz direktnu interakciju sa ligandom, ili kroz alosterne efekte²¹⁴. Radi sticanja atomističkog uvida u opažene efekte WDR5:*cis*-CS-VIP 8 interakcije, izvršene su molekulske dinamičke (MD) simulacije na četiri različita sistema: Sistem A: *apo* WDR5; Sistem B: WDR5:*cis*-CS-VIP 8; Sistem C: MLL1 peptid-WDR5-RbBP5 peptid; Sistem D: *cis*-CS-VIP 8-WDR5-RbBP5 peptid.

Sistem A, *apo* WDR5, je proučavan kako bi se razumela inherentna dinamika WDR5 van konteksta vezivanja liganda. Sistem B, WDR5:*cis*-CS-VIP 8, je proučavan kako bi se istražio uticaj vezivanja liganda na dinamiku WDR5. Sistem C, MLL1 peptid-WDR5-RbBP5 peptid, je proučavan kako bi se razumela dinamika WDR5 u kontekstu interakcije sa svojim prirodnim vezivnim partnerima. Sistem D, *cis*-CS-VIP 8-WDR5-RbBP5 peptid, je proučavan kako bi se istražio uticaj vezivanja liganda na dinamiku celog kompleksa WDR5-RbBP5.

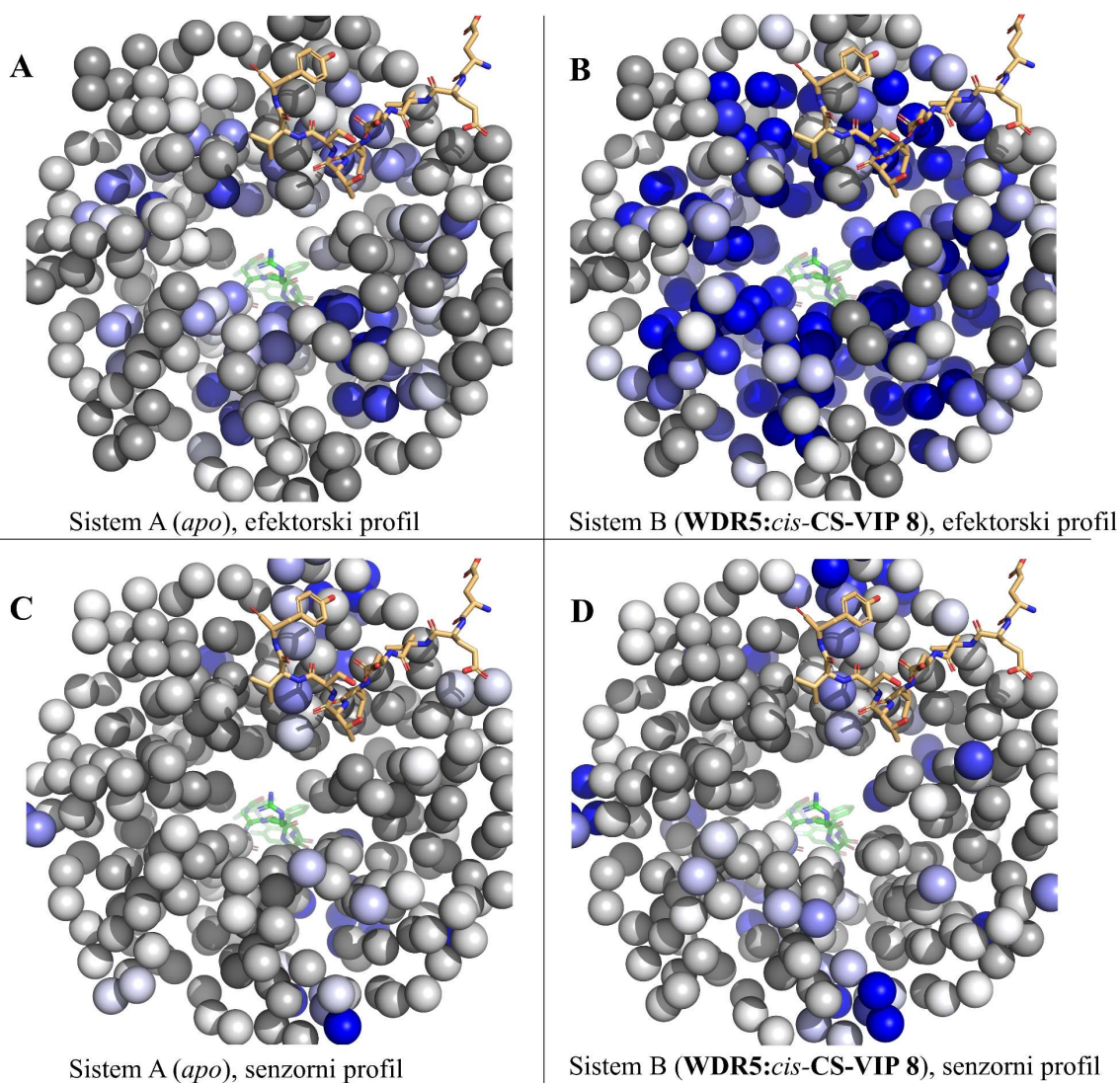
Eksperimenti praćenja stope izmene vodonika i deuterijuma unutar kompleksa MLL1 uzrokovanih prisustvom *cis*-CS-VIP 8 ukazali su na značajno izmenjene profile za gotovo sve komponente MLL1 kompleksa.²¹⁴ Polipeptidna sekvenca MLL1 koja interaguje sa WIN mestom WDR5, tzv. Win peptid, je pokazala povećane stope zamene vodonika/deuterijuma, kao i samo jezgro MLL1 SET domena, što ukazuje na gubitak interakcije sa WDR5 i RbBP5/Ash2L uzrokovano prisustvom *cis*-CS-VIP 8²¹⁴. Odgovarajuće, aminokiseline 336-354 RbBP5 koje povezuju SET i Ash2L takođe su pokazale povećane stope zamene vodonika/deuterijuma, što ukazuje na narušavanje strukture ovog dela kompleksa usled vezivanja *cis*-CS-VIP 8. Značajno je da *cis*-CS-VIP 8 indukuje povećanu vodonik/deuterijum izmenu u ostacima WDR5 220-230, koji su udaljeni od mesta vezivanja liganda i poklapaju se sa RbBP5 mestom vezivanja, odnosno WBM vezivnim mestom.



Slika 4.33. Rezultati SASA analize.

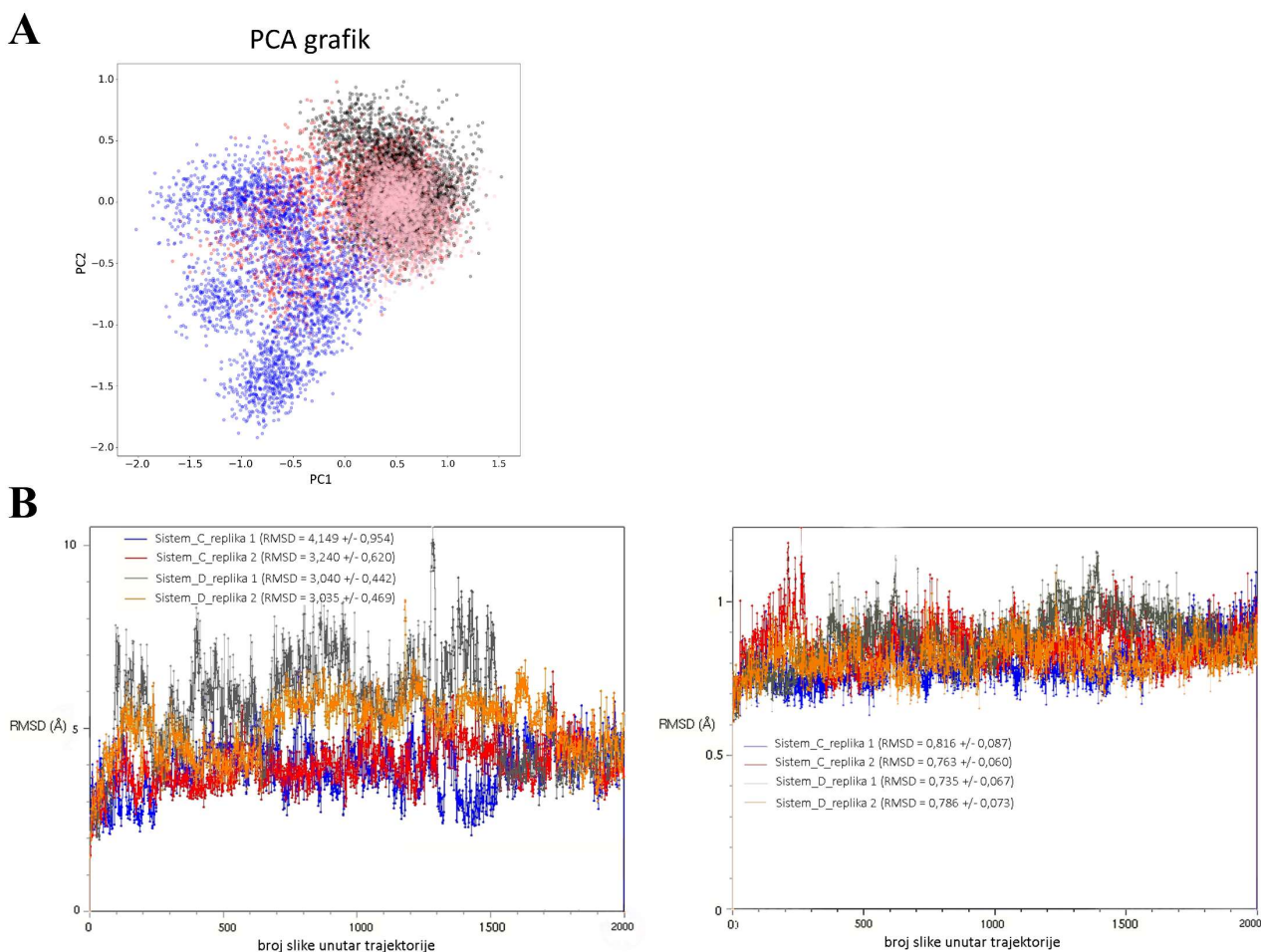
Prethodno izvedeni HDX eksperimenti su ukazali na značaj WDR5 ostataka 220-230 (WBM mesto) za dinamičko ponašanje WDR5, kao i celog MLL1 kompleksa. Kako bi se potvrdio potencijalni mehanizam deuterijumske izmene u WDR5 strukturama van vezivnog mesta liganda, razlike u dostupnim površinama molekula solventa (eng. *Solvent accessible surface area* – SASA) izračunate su za vodonike iz peptidnih veza unutar WDR5 za izučavane sisteme A i B. Rezultati SASA analize su bili usaglašeni sa HDX eksperimentima²¹⁴ i ukazali su na gotovo identične regione unutar WDR5 (Slika 4.33) što je dodatno ukazalo na značajnu izmenu u konformacionoj dinamici WDR5 prisustvom *cis*-CS-VIP 8.

S obzirom da su HDX eksperimenti i SASA analiza MD trajektorija ukazali na potencijalne promene u konformacionoj dinamici delova WDR5 udaljenih od mesta vezivanja *cis*-CS-VIP 8, pretpostavljen je alosterni efekat inhibitora i dalje se pristupilo detaljnijoj analizi mehanizma alosterije.



Slika 4.34. Profil efektorskih rezidua (A i C) i senzorskih rezidua (B i D) izražen kao B-faktori na $C\alpha$ atomima za Sistem A i Sistem B. RbBP5 peptid (naranžasta) i *cis*-CS-VIP 8 (zeleni) prikazani su na svim slikama kao indikatori položaja vezivnih mesta za RbBP5 i MLL1. Skala boja sive-belo-plava označava ostatak sa niskim vrednostima B-faktora kao siv, a ostatak sa visokim vrednostima B-faktora kao plav. Raspon vrednosti za kodiranje boja je bio 0-0,2.

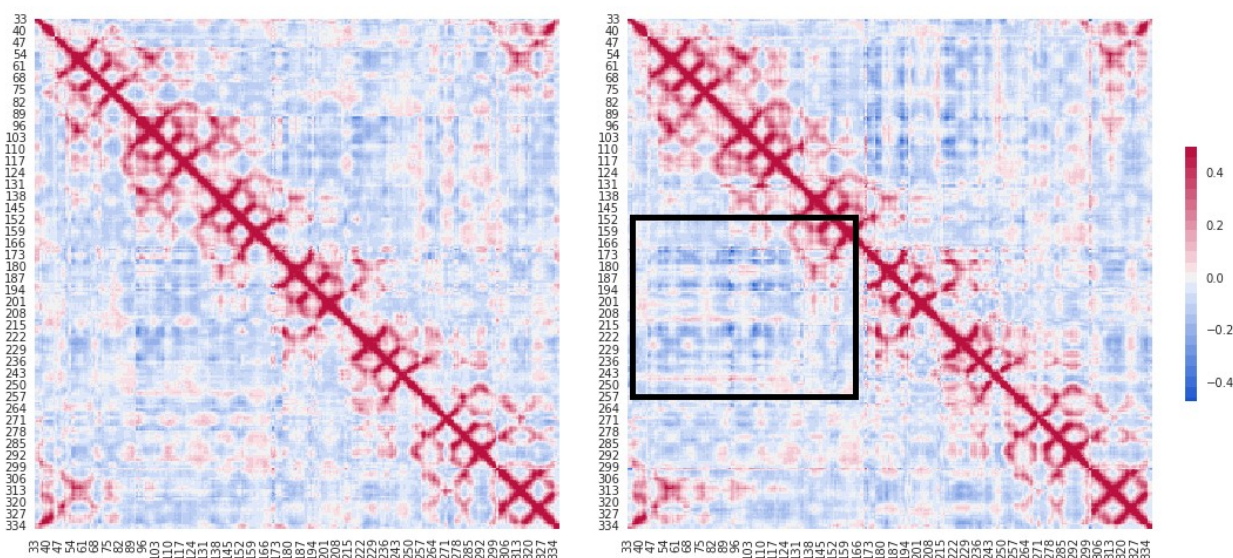
Analiza skeniranja perturbacionog odgovora (eng. *Perturbation response scanning* – PRS) je noviji alat u analizi mehanizama transdukcije alosternih signala unutar strukture proteina³³⁰. PRS metodom analizira se uloga svake od aminokiselina unutar izučavanog proteina na transdukciju signala izazvanog odgovarajućom perturbacijom (npr. vezivanje liganda). Rezultati PRS-a se obično tumače preko profila senzorskih i efektorskih aminokiselinskih ostataka. Efektorski ostaci su oni čija perturbacija rezultira "konformacionim" promenama na udaljenim mestima proteina (senzorski ostaci). Matrice kovarijansi generisane analizom trajektorija MD simulacija upotrebljene su za PRS analizu. Rezultati PRS analize izvršene na sistemima A (*apo* WDR5, poglavlje 3.2.2) i B (WDR5:*cis*-CS-VIP 8) i izraženi kroz B-faktore pokazali su da se većina efektor mapira na MLL1 vezno mesto i centralni kanal WDR5, dok se većina senzora mapira na i oko RbBP5 vezujućeg (WBM) mesta. Međusobno upoređivanje profila efektor (Slika 4.34 A i C) ukazuje na to da se snažna komunikacija u sistemu B ostvaruje kroz centralni kanal WDR5 kao rezultat vezivanja liganda. Profili senzora ukazuju na blage promene oko RbBP5 vezujućeg mesta (Slika 4.34 D i C). Ovi rezultati su bili prvi pokazatelj alosterne komunikacije između MLL1 i RbBP5 mesta na WDR5.



Slika 4.35. Rezultati PCA i RMSD analize. A – Rezultati PCA analize (crvena - Sistem A, plava - Sistem B, pink - Sistem C, crna - Sistem D); B – RMSD vrednosti za Sisteme C i D izračunate za sve atome WDR5 (levo) i za ostatke RbBP5 sekvence koji interaguju sa ostatcima 220-230 WDR5 (WBM mesto) (desno). Sistemi C i D su simulirani u po dve replike, od po 200 ns.

Rezultati analize glavnih komponenti atomskih pozicija atoma peptidnog skeleta dobijenih tokom MD simulacija, ukazuju na različito konformaciono ponašanje simuliranih sistema. Najveća razlika je primećena između sistema A i B (A: *apo* WDR5, B: WDR5:*cis*-CS-VIP 8) i sistema C i D (C: MLL1 peptid-WDR5-RbBP5 peptid; D: *cis*-CS-VIP 8-WDR5-RbBP5 peptid). Dok su sistemi C i D dobro lokalizovani na ravni PC1 i PC2 projekcija, što ukazuje na konformaciono stabilnije sisteme zarobljene u nižim energetskim minimumima, sistemi A i B su više delokalizovani i samo se delimično preklapaju u projekcijama konformacionog prostora (Slika 4.35A). Analizom RMSD i vizualizacijom trajektora za sisteme C i D, primećeno je delimično odvezivanje RbBP5 peptida u sistemu D, što ukazuje na razlike u konformacionom ponašanju u WDR5 izazvanom različitim ligandima prisutnim u WIN vezivnom mestu WDR5 (Slika 4.35B).

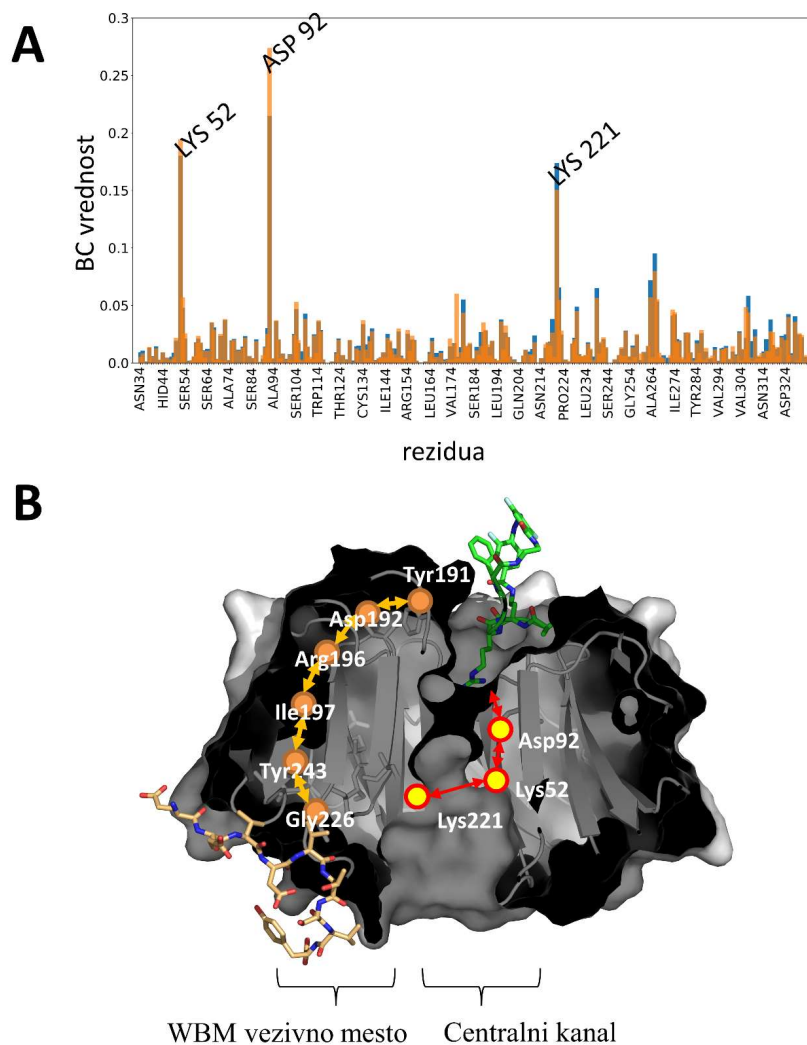
Poređenjem matrica ukrštene korelacije atomskih pozicija sistema C i D, u sistemu D oko ostataka 220-230 (RbBP5 vezujućeg mesta) uočeni su anti-korelisani pokreti što dodatno ukazuju na alosterne efekte *cis*-CS-VIP 8. Sveukupno, PCA i konformacionom analizom utvrđen je uticaj alosternih efekata *cis*-CS-VIP 8 koji dovode do različitog dinamičkog ponašanja WDR5:*cis*-CS-VIP 8 kompleksa u odnosu na alosterne efekte MLL1 peptida (Slika 4.36). Ovi rezultati ukazuju da inhibicija MLL1 kompleksa putem kompeticije sa MLL1 na WDR5 (u okviru WIN mesta) nije jedini mehanizam uključen u disocijaciju komponenti MLL1 kompleksa, već se i alosterne efekti liganda mogu smatrati značajnim u pospešivanju disocijacije što opravdava primenu MD simulacija u dizajnu novih inhibitora.



Slika 4.36. Ukršteni korelacioni matrice za sistem C (levo) i sistem D (desno). Bojenje označava anti-korelirane pokrete u plavoj boji. Najizraženiji anti-korelirani pokreti su označeni crnim kvadratom. Brojevi na x i y-osi označavaju identifikacione brojeve rezidua.

Analiza dinamičke interakcione mreže rezidua (eng. *Dynamical residue interaction network* – RIN) upotrebljena je kako bi se dublje razumelo konformaciono ponašanje uočeno tokom MD simulacija (opisano iznad) i kako bi se dodatno potvrdili rezultati PRS analize. Rezultati ove analize su rasvetlili najznačajnije komunikacione puteve između dva vezivna mesta WDR5 (WIN i WBM mesto). Mera uzajamne centralizovanosti (eng. *Betweenness centrality* – BC) upotrebljena je za detektovanje čvorova (aminokiselinskih rezidua) koji imaju veliki uticaj na protok informacija unutar interakcione mreže. U svim simuliranim sistemima, rezidue Lys52, Asp92 i Lys221 su identifikovane kao najvažniji komunikacioni čvorovi u WDR5 (Slika 4.37). Zbog njihove prostorne lokalizacije (centralnog kanala WDR5 (Slika 4.37)), ovi čvorovi su predloženi kao posrednici u alosternoj komunikaciji između WIN vezivnog mesta (MLL1 peptid/*cis*-CS-VIP 8) i WBM vezivnog mesta (RbBP5 komponenta MLL1 kompleksa) na WDR5. Prema analizi najkraćeg puta, direktna interakcija aminokiseline Asp92 sa argininom iz *cis*-CS-VIP 8 ili MLL1 peptida je važna za pokretanje komunikacije kroz centralni kanal. Signal se zatim propagira kroz soni most između Asp92 i Lys52 i dalje se prenosi na drugu stranu centralnog kanala (Slika 4.37). Prenos signala na drugu stranu centralnog kanala i komunikacija sa WBM mestom (aminokiseline 220-230 i 240-250) olakšava se formiranjem vodom posredovanih vodoničnih veza između Lys52 i Lys221. Dodatna analiza najkraćeg puta otkrila je alternativni put koji povezuje aminokiselinu Tyr191 sa vezivanjem RbBP5 preko površine WDR5 (Slika 4.37).

U cilju daljeg razjašnjavanja mehanizama uticaja prisustva ili odsustva inhibitora i/ili RbBP5 na strukturu mreža interakcija, izvršena je uporedna analiza preusmeravanja komunikacionih puteva (eng. *rewiring*) (Prilog 7.12). Upoređujući sisteme A (*apo* WDR5) i B (WDR5:*cis*-CS-VIP 8), svi čvorovi sa najvećim BC su detektovani kao značajno promenjeni, što se kvantifikuje većim vrednostima Dn-skorova (metrika preusmeravanja, poglavlje 3.2.2). Naime, vezivanje liganda izazvalo je značajno preusmeravanje mreže interakcija, što se manifestuje kroz velike Dn-skorove. Dodatna analiza i vizualizacija trajektorija potvrđuje da u *apo* sistemu postoji značajna razlika/promena u protoku informacija u poređenju sa sistemima sa ligandima, izazvana delimično neusaglašenim pokretima Asp92 rezidue, koji bi inače bili stabilizovani direktnom interakcijom sa ligandima.



Slika 4.37. Rezultati analize alosterne komunikacije. A – BC analiza za MLL1-Win-WDR5-RbBP5 kompleks (narandžasta) i *cis*-CS-VIP 8-WDR5-RbBP5 kompleks (plava); B – Uspostavljeni model alosterne komunikacije kroz WDR5.

Upoređujući sisteme C i D, najveće preusmeravanje detektovano je za ostatke oko MLL1 vezivnog mesta (Prilog 7.12), što ukazuje da razlike u interakcijama CS-VIP 8 i MLL1 peptida uzrokuju razlike u konformacionom ponašanju WDR5 opisane kroz konformacionu analizu predstavljenu u tekstu iznad. Zanimljivo je da detektovano preusmeravanje Asp92 između sistema C i D, kao jedne od najvažnijih aminokiselinskih rezidua za prenos alosternog signala kroz centralni kanal WDR5, može objasniti uočeno delimično odvajanje RbBP5 tokom simulacije sistema D.

Dodatna analiza preusmeravanja mreže ukazala je da su aminokiseline uključene u put koji povezuje aminokiselinu Tyr191 sa mestom vezivanja RbBP5 (Slika 4.37) (Asp192, Arg196) takođe pogodene prisustvom ili odsustvom različitih liganda na mestu vezivanja MLL1 (Prilog 7.12). Značaj Tyr191 u enzimskoj aktivnosti MLL1 kompleksa je prethodno eksperimentalno utvrđena studijama mutageneze³³¹. Uzeti zajedno, rezultati simulacija ukazuju na to da suptilni pomaci u obrascima interakcija i preusmeravanje mreže interakcija aminokiselina WDR5 igraju glavnu ulogu u alosternoj komunikaciji između WIN i WBM mesta vezivanja dok pojava velikih konformacionih promena unutar samog WDR5 izostaje (što je u skladu sa prijavljenim kristalografskim studijama). Dodatno, rezultati simulacija ističu značaj komunikacije unutar centralne pore WDR5 inicirane prisustvom arginina u WIN vezivnom mestu za stabilizaciju celokupnog MLL1 kompleksa.

4.5. Bioinformatička analiza novih ciljnih mesta za razvoj dualnih inhibitora histon-deacetilaza kao potencijalnih terapeutika u tretmanu adenokarcinoma pankreasa

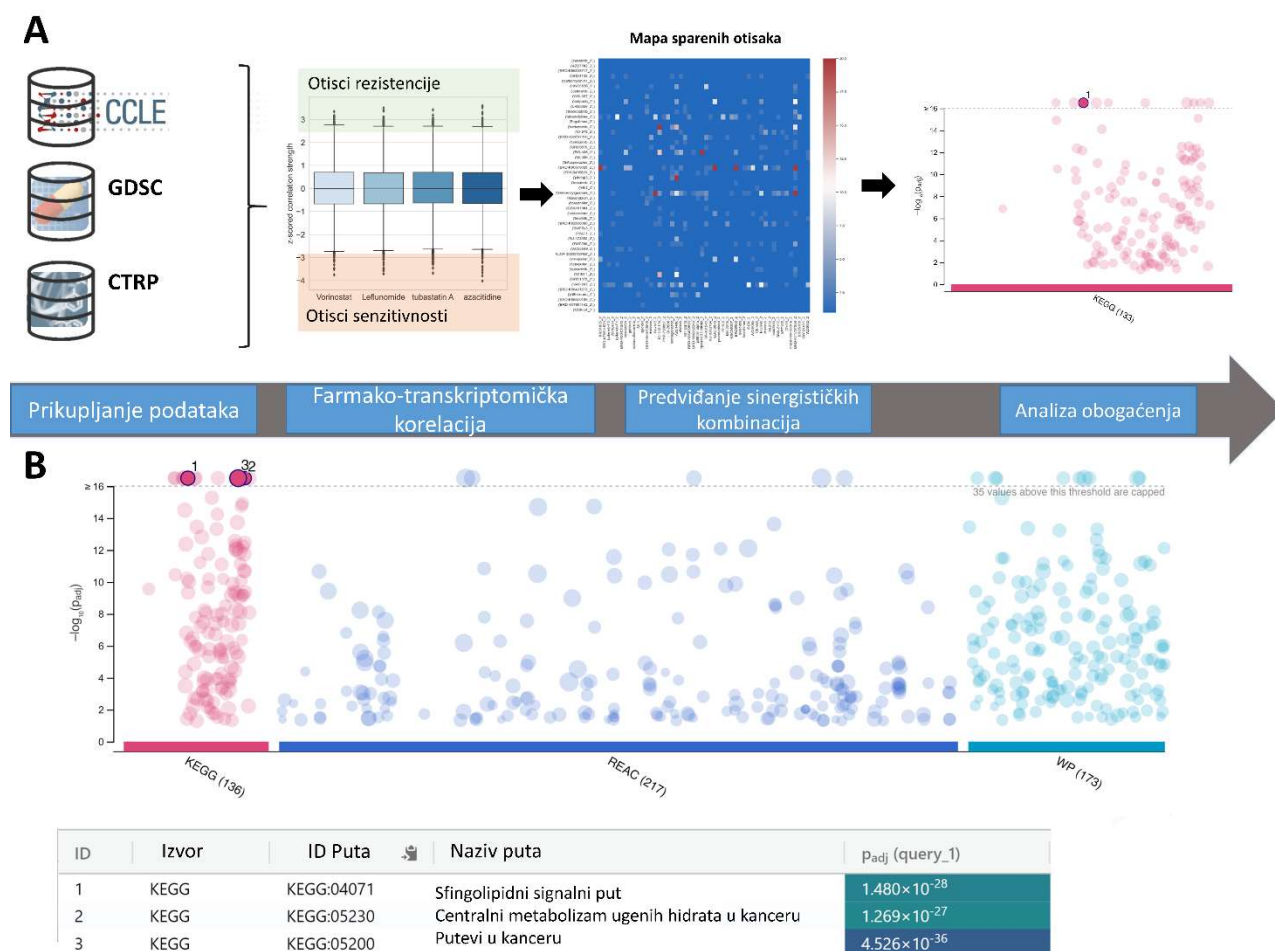
Duktalni adenokarcinom pankreasa (eng. *Pancreatic ductal adenocarcinoma* – PDAC) se smatra jednim od najagresivnijih i najsmrtonosnijih maligniteta. Rana dediferencijacija i veliki metastatski potencijal značajno otežavaju rano dijagnostikovanje bolesti.^{332–334} Iako hirurgija praćena adjuvantnom hemioterapijom pruža izgleda za dugoročno preživljavanje, kod preko 80% pacijenata sa PDAC-om bolest se dijagnostikuje u terminalnim fazama kada su već prisutne udaljene metastaze^{335–337}. Pored toga, većina pacijenata koji su podvrgnuti resekciji doživljava povratak bolesti u roku od godinu dana, što dodatno ističe značaj hemioterapije u lečenju PDAC³³⁷. S obzirom na veoma skromno poboljšanje petogodišnje stope preživljavanja od 3% do 11% od 1970. godine do danas, razvoj efikasnijih nehirurških terapijskih pristupa u lečenju PDAC-a predstavlja suštinsku potrebu.

Gusta stroma zajedno sa brojnim genetskim i epigenetskim promenama doprinosi kompleksnim mehanizmima rezistencije PDAC na dostupne hemioterapeutike. Inspirisana mnogim epigenetskim promenama u PDAC-u, pretklinička istraživanja koja proučavaju interakciju epigenetskih terapeutika sa epigenetskim mehanizmima koji leže u osnovi PDAC-a dala su obećavajuće rezultate. Ipak, sve više dokaza ukazuje da epigenetski terapeutici pružaju merljive koristi u PDAC-u samo u kombinovanoj terapiji^{336,338}. U okviru epigenetskih grupa lekova, inhibitori histon deacetilaza (HDACi) predstavljaju jedne od najperspektivnijih terapeutika u terapiji PDAC-a³³⁸. Trenutno postoji četiri HDACi odobrena od strane Američke agencije za hranu i lekove (FDA). Uprkos obećavajućim pretkliničkim rezultatima, klinička ispitivanja koja istražuju efikasnost HDACi kao pojedinačnih agenasa u solidnim tumorima dala su prilično razočaravajuće rezultate³³⁹ zbog čega je kombinacija HDACi sa drugim hemioterapeuticima predložena kao jedan od najperspektivnijih pristupa u budućem razvoju HDACi-bazirane hemioterapije^{338,340}.

U cilju pronalaženja novih ciljnih mesta dejstva za razvoj inhibitora histon-deacetilaza sa dualnom aktivnošću razvijen je protokol za ispitivanje farmako-transkriptomčke veze između bazalne genske ekspresije i hemijske osetljivosti ćelijskih linija karcinoma pankreasa korišćenjem sistematske korelacione analize velikih javno dostupnih setova podataka. Iako do danas izostaju klinička ispitivanja sirtuinskih inhibitora, bioinformatičkom analizom je bila obuhvaćena cela familija histon-deacetilaza.

Povezivanje podataka o hemijskoj osetljivosti sa bazalnom genskom ekspresijom prethodno je pokazano kao vredna metoda za otkrivanje mehanizama delovanja bioaktivnih jedinjenja²²², ili za definisanje transkriptomskih otisaka odgovora ćelijskih linija karcinoma na terapiju za ograničen broj HDAC inhibitora²²³. U ovom istraživanju su podaci o hemijskoj osetljivosti preuzeti iz javno dostupnih baza poput genomičke baze osetljivosti malignih ćelija na lekove „GDSC“ (eng. *Genomics of Drug Sensitivity in Cancer* – GDSC) i portala terapijskog odgovora malignih ćelija „CTRP“ (eng. *Cancer Therapeutics Response Portal* – CTRP) za 864 malih molekula na do 38 ćelijskih linija poreklom iz karcinoma pankreasa. Profili osetljivosti su zatim korelisani sa bazalnom genskom ekspresijom 19221 gena dobijenih iz Enciklopedije tumorskih ćelijskih linija (eng. *Cancer Cell Line Encyclopedia* – CCLE)³⁰³. Izračunati su koeficijenti *Pearson*-ove korelacije između vrednosti površine ispod krive i podataka o bazalnoj ekspresiji (izraženi kao log₂RSEM vrednosti) za svaki transkript. Matrica koeficijenata *Pearson*-ove korelacije dalje je normalizovana korišćenjem *Fisher*-ove *z*-transformacije kako bi se prilagodila varijacijama u broju testiranih ćelijskih linija za svaku analiziranu molekulu. Uopšteno govoreći, transkripti negativno korelisani sa podacima o hemijskoj osetljivosti mogu se pripisati osetljivosti ćelijskih linija karcinoma pankreasa prema odgovarajućim malim molekulama, dok se pozitivno korelisani transkripti mogu pripisati rezistenciji ćelijskih linija. U skladu s tim, transkripti koji odgovaraju *Tukey*-evim ekstremima za svako bioaktivno jedinjenje dalje su izdvojeni i sačuvani kao transkriptomski otisci osetljivosti ili rezistencije.

Formirani transkriptomski otisci su dalje upotrebljeni za predviđanje novih sinergističkih kombinacija bioaktivnih jedinjenja. Pod pretpostavkom da su geni iz otisaka rezistencije direktno uključeni u mehanizme tumorskih ćelija koji omogućavaju izbegavanje smrti ćelija pod uticajem leka, dok su geni iz otisaka osetljivosti direktno uključeni u mehanizme koji senzitivizuju tumorske ćelije na terapiju, pretpostavljeno je da nivo preklapanja između ovih otisaka može biti pokazatelj sinergizma između malih molekula i može identifikovati nove ciljane molekule za dizajn dualnih HDAC inhibitora u terapiji PDAC. Drugim rečima, pretpostavljeno je se da se sinergistički efekti između dva bioaktivna jedinjenja javljaju kada prvo jedinjenje koji povećava osetljivost tumorske ćelije koristi isti, ili sličan skup gena poput onih koji su uključeni u razvoj rezistencije na drugo jedinjenje. Opšti pristup studije prikazan je na Slici 4.38.



Slika 4.38. Prikaz osnovnog protokola bioinformatičke analize (A) i rezultata analize obogaćenja izdvojenih signalnih puteva (B).

U cilju dalje identifikacije novih sinergističkih kombinacija HDACi i preostalih bioaktivnih jedinjenja, izvršeno je prebrojavanje preklapanja transkripata u skupu korelacionih podataka, pri čemu su uzeti u obzir samo transkripti sa značajnom korelacijom (videti poglavlje 3.3). Ukoliko su se transkripti sa značajnom pozitivnom korelacijom odgovora prvog bioaktivnog jedinjenja preklapali sa transkriptima povezanim sa negativnom korelacijom odgovora drugog bioaktivnog jedinjenja, broj preklapanja je prebrojan i sačuvana (Tabela 4.8 – Broj +/- i Broj -/+). Skorovi za sve moguće kombinacije HDAC inhibitora i ostalih malih molekula su izračunati, a verovatnoća svake kombinacije je procenjena kao broj svih mogućih preklapanja između značajno korelisanih transkripata (nazvan „*Final_Syn_Score*“). Konačni rezultat obuhvata 29233 moguće kombinacije između HDACi i preostalih bioaktivnih jedinjenja. Celokupna tabela je dostupna u okviru publikacije³⁴¹.

Korišćenjem predloženog protokola, identifikovano je nekoliko poznatih sinergističkih partnera HDACi, što doprinosi validnosti dizajniranog protokola. Na primer, nedavno je opisana sinergistička interakcija između inhibitora DNMT1 i pan-HDACi na ćelijama MIA PaCa2 adenokarcinoma pankreasa³⁴². Sorafenib je jedan od inhibitora kinaza (BRAF; FLT3; KDR; RAF1) sa prekliničkim podacima o sinergizmu sa HDACi, a ova kombinacija se trenutno ispituje u kliničkim istraživanjima PDAC-a^{338,343}. Još neki pozitivni primeri uključuju kombinaciju HDAC inhibitora sa inhibitorom fosfoinozimid-3 kinaze (PI3K)³⁴⁴, mTOR (eng. *Mammalian target of rapamycin* – mTOR) inhibitorom³⁴⁵, BET inhibitorima (BRD4 u Tabeli 4.8)³⁴⁶, i kombinaciju sa gemcitabinom³⁴⁰ (Tabela 4.8). Treba napomenuti da nisu svi inhibitori navedenih farmakoloških ciljnih mesta imali visok „*Final_Syn_Score*“, što se može pripisati uglavnom nepoznatim sporednim (eng. *off-target*) efektima bioaktivnih jedinjenja. Interesantno je uočiti da je protokol izdvojio inhibitore ciklin-zavisne kinaze 9 (CDK9) i mTOR inhibitore, kao potencijalne sinergističke partnere selektivnim SIRT2, ili SIRT1 inhibitorima (Tabela 4.8). Iako za SIRT inhibitore još uvek ne postoji dovoljno eksperimentalnih podataka o značaju u tretmanu PDAC, oba ciljna mesta su već razmatrana kao nova obećavajuća mesta za terapiju što otvara vrata budućim istraživanjima^{347,348}.

U cilju identifikacije mogućih signalnih puteva koji bi mogli biti ciljani kako bi se postigao sinergijski efekat sa HDAC inhibitorima u PDAC-u, izvršena je analiza obogaćenja. Analizom obogaćenja (eng. *Enrichment analysis*) bila su obuhvaćena sva ciljna mesta predviđena kao potencijalni sinergistički interaktori HDAC inhibitora. Analiza obogaćenja je pokazala sfingolipidni signalni put (kako je definisano u KEGG bazi podataka)^{349,350} kao jedan od najistaknutijih puteva uključenih u predviđene sinergizme između analiziranih bioaktivnih jedinjenja (Slika 4.38). Sfingolipidni signalni put obuhvata nekoliko enzima koji su uključeni u metabolizam sfingolipida (sfingozin kinaza (SPHK), ceramid kinaza (CERK), ceramid sintaza (CERS), alkalna ceramidaza (ACER), kiselinska ceramidaza (ASAH) itd.), kao i receptore za sfingozin-1-fosfat u membrani (S₁PR₁₋₅) i njihove nishodne efektorske Rho/ROCK, PI₃K/Akt i MAPK puteve (kako je definisano kroz analizirane KEGG i Wiki Pathway baze podataka³⁵¹). Interesantno je napomenuti da trenutni literaturni podaci ne prepoznaju nijedan od elemenata sfingolipidnog signalnog puta (osim nishodnih puteva aktiviranih različitim signalima) kao potencijalna nova ciljna mesta za postizanje sinergizma sa HDACi.

U cilju eksperimentalne validacije protokola, analizirani su sinergizmi između novih HDAC inhibitora i modulatora sfingolipidnog signalnog puta dostupnih u laboratoriji tokom sprovođenja istraživanja. Za ovu analizu upotrebljen je skup nedavno razvijenih HDAC inhibitora: jedinjenje 6b (IC₅₀ (HDAC1) = 4.73 μM, IC₅₀ (HDAC3) = 1.86 μM, IC₅₀ (HDAC6) = 0.19 μM, IC₅₀ (HDAC8) = 2.44 μM); jedinjenje 8b (IC₅₀ (HDAC1) = 0.41 μM, IC₅₀ (HDAC3) = 0.21 μM, IC₅₀ (HDAC6) = 0.12 μM, IC₅₀ (HDAC8) = 0.24 μM); jedinjenje 9b (IC₅₀ (HDAC1) = 1.47 μM, IC₅₀ (HDAC3) = 3.47 μM, IC₅₀ (HDAC6) = 0.03 μM, IC₅₀ (HDAC8) = 0.71 μM)³⁵².

Tabela 4.8. Prikaz izdvojenih sinergizama.

Molekul 1 (HDACi)	Molekul 2	Ciljna mesta molekula 1	Ciljna mesta molekula 2	Broj +/- ^a	Broj -/+ ^b	Final_Syn_Score ^c
tacedinalin	RG-108	HDAC1;HDAC2;HDAC3; HDAC6;HDAC8	DNMT1	37	403	440 (0,3%)
Merck60	Sorafenib	HDAC1;HDAC2	BRAF;FLT3; KDR;RAF1	14	61	75 (5%)
tacedinalin	PI-103	HDAC1;HDAC2;HDAC3; HDAC6;HDAC8	PI3K α , DAPK3, CLK4, PIM3, HIPK2	23	243	266 (1%)
tacedinalin	KU-0063794	HDAC1;HDAC2;HDAC3; HDAC6;HDAC8	MTOR	189	425	614 (0,1%)
BRD-K85133207	AZD5153	HDAC1	BRD4	36	47	83 (4,6%)
PCI-34051	gemcitabin	HDAC8, HDAC6, HDAC1	CMPK1;RRM1;TYMS	32	34	66 (5,7%)
PCI-34051	GSK429286A	HDAC8, HDAC6, HDAC1	ROCK1, ROCK2	65	79	144 (2,3%)
BRD-K51490254	SKI-II	HDAC6;HDAC8	SPHK1	12	63	75 (5%)
PCI-34051	ingolimod	HDAC8, HDAC6, HDAC1	S1PR1	10	17	27 (12,7%)
AGK-2	BRD-K30748066	SIRT2	CDK9	68	218	286 (0,92%)
EX-527	temsirolimus	SIRT1	MTOR	42	100	142 (2,36%)

^a Broj sparenih pozitivno koreliranih (molekul 1) i negativno koreliranih (molekul 2) transkripta; ^b Broj sparenih negativno koreliranih (molekul 1) i pozitivno koreliranih (molekul 2) transkripta; ^c Brojevi u zagradama ukazuju na poziciju datog rešenja unutar rangirane liste.

Jedna od kombinacija među 2,3% predviđenih sinergizama sortiranih na osnovu „Final_Syn_Score” skora bila je kombinacija inhibitora Rho-udružene protein kinaze (eng. *Rho-associated protein kinase* – ROCK) 1/2 i pan-HDACi (Tabela 4.8). Interesantno je da je inhibicija ROCK-a, pored efekata na invaziju i rast tumora, prethodno opisana kao vredna „pripremna“ strategija za hemioterapiju PDAC-a^{353,354}. ROCK je takođe identifikovan kao jedan od efektoru u nizu sfingolipidne signalizacije S1P^{350,355}. U eksperimentalnoj analizi upotrebljen je inhibitor ROCK1/2 RKI-1447 (IC₅₀ (ROCK1) = 14.5 nM i IC₅₀ (ROCK2) = 6.2 nM). Fingolimod – još jedno jedinjenje koje je predviđeno kao sinergistički partner pan-HDACi, predstavlja agonistu i funkcionalnog antagonistu podtipa 1 receptora za S1P (S1PR1). Fingolimod se pokazao kao potentan antitumorski agens u određenim modelima PDAC-a, što se može delimično pripisati i umerenim inhibitorynim efektima na klasu I HDAC-a^{356,357}. Osim receptora za S1P, drugi elementi sfingolipidnog signalnog puta prepoznati su kao sporedna ciljna mesta (*off-target*) fingolimoda, uključujući ceramid sintazu 2 (CERS2), S1P liazu i SPHK1. Sveukupno, fingolimod predstavlja interesantnu hemijsku probu za ispitivanje potencijala sinergizma između HDACi i modulacije sfingolipidnog signalnog puta³⁵⁸. Važno je napomenuti da je par fingolimod:HDACi rangiran među prvih 13% predikcija, što ukazuje na umeren sinergizam.

Pri eksperimentalnoj proceni citotoksičnih efekata kombinacija bioaktivnih jedinjenja, korišćen je Chou-Talalay model³⁰⁵, koji zahteva da supstance budu primenjeni u fiksnom odnosu doza. MIA PaCa-2 i Panc-1 ćelije su tretirane kombinacijom sintetisanih HDAC inhibitora i RKI-1447, ili HDAC inhibitora i fingolimoda u matrici doza 4 × 4, a rezultatna preživljavanja ćelija su izmerena pomoću MTT testa. Analiza izoblograma za kombinaciju bioaktivnih jedinjenja pri niskim koncentracijama i koncentracijama do IC₅₀ vrednosti pokazala je sinergizam za kombinaciju jedinjenja 6b i RKI-1447, kao i jedinjenja 9b i RKI-1447 na MIA PaCa-2 ćelijama (Tabela 4.9).

Sinergistički efekti kombinovanog tretmana sa jedinjenjem 8b i RKI-1447 na preživljavanje ćelija su primećeni samo pri najvišoj testiranoj koncentraciji (Tabela 4.9). Kada su u pitanju Panc-1 ćelije,

analiza izoblograma kombinovanog tretmana pokazala je blagi sinergizam samo za kombinaciju jedinjenja 9b i RKI-1447 pri najnižim testiranim koncentracijama (Tabela 4.9). Slično rezultatima za sinergizam između HDACi i RKI-1447, sinergizam između HDAC inhibitora i fingolimoda takođe je primećen pretežno kod jedinjenja 6b i 9b (Tabela 4.10).

U skladu sa predviđanjima, sinergizmi između HDAC inhibitora i fingolimoda su umereni i uglavnom su primećeni za najviše testirane koncentracije. Iako su utvrđeni antagonizmi za neke od testiranih parova koncentracija, ovi antagonizmi su manje zabrinjavajući jer je većina njih detektovani za niske frakcije preživelih ćelija (treći redovi Tabela 4.9 i 4.10 za svaki par).

Tabela 4.9. Analiza sinergizama za kombinaciju novih HDAC inhibitora i RKI-1447.

	MIA PaCa-2 ćelijska linija				Panc-1 ćelijska linija			
	Koncentracija jedinjenja ^a				Koncentracija jedinjenja ^a			
	I	II	III	IV	I	II	III	IV
HDACi (6b)	48,60 ± 7,45	13,58 ± 2,33	4,80 ± 1,01	2,91 ± 0,87	23,61 ± 3,26	1,39 ± 0,55	0,47 ± 0,21	0,5 ± 0,2
ROCKi (RKI-1447)	64,18 ± 8,12	40,49 ± 5,55	37,49 ± 3,02	30,84 ± 7,33	62,76 ± 11,23	33,89 ± 8,92	22,02 ± 7,22	18,11 ± 5,90
6b + RKI-1447	83,28 ± 10,34	55,48 ± 7,51	37,69 ± 5,38	38,31 ± 4,56	69,79 ± 10,45	48,51 ± 6,34	24,98 ± 3,13	8,92 ± 4,21
Kombinatorijski indeks (CI)	0,445	0,981	1,253	0,605	0,992	1,050	1,334	2,117
Interakcija 6b + RKI-1447 ^b	+++	+	--	++	±	±	--	---
HDACi (8b)	27,30 ± 4,44	1,83 ± 0,70	0,94 ± 0,45	0,57 ± 0,19	49,78 ± 6,77	38,21 ± 5,89	17,13 ± 2,11	0,1 ± 0,05
ROCKi (RKI-1447)	62,94 ± 10,11	42,93 ± 6,99	36,07 ± 7,51	35,46 ± 4,09	58,78 ± 5,55	30,77 ± 3,45	17,27 ± 2,28	7,41 ± 2,56
8b + RKI-1447	73,02 ± 7,44	45,32 ± 5,14	31,14 ± 5,21	28,91 ± 2,34	63,94 ± 6,33	42,96 ± 5,79	21,99 ± 3,45	6,79 ± 3,06
Kombinatorijski indeks (CI)	0,333	1,564	2,469	1,505	3,133	1,433	1,266	1,474
Interakcija 8b + RKI-1447 ^b	+++	---	---	---	---	--	--	---
HDACi (9b)	44,65 ± 3,39	18,25 ± 2,22	1,71 ± 0,67	0,46 ± 0,20	45,18 ± 4,55	23,14 ± 7,42	7,56 ± 3,71	1,12 ± 0,89
ROCKi (RKI-1447)	61,64 ± 7,78	43,28 ± 5,38	35,02 ± 3,25	29,16 ± 2,16	64,82 ± 6,41	37,92 ± 5,34	19,76 ± 2,30	9,41 ± 1,06
9b + RKI-1447	78,03 ± 7,18	64,37 ± 6,78	41,91 ± 5,98	33,32 ± 4,21	64,68 ± 5,01	51,26 ± 7,20	33,14 ± 6,10	18,01 ± 4,02
Kombinatorijski indeks (CI)	0,513	0,577	0,964	0,803	1,705	1,207	0,975	0,823
Interakcija 9b + RKI-1447 ^b	+++	+++	±	++	---	--	±	+

^a Koncentracije su prikazane u silaznom nizu (I je najviša. IV je najniža). ^b Tumačenje: (+++++) Veoma jak sinergizam; (++++) Sinergizam; (++) Umeren sinergizam; (+) Slab sinergizam; (±) Približno aditivni; (-) Slab antagonizam; (--) Umeren antagonizam; (---) Antagonizam; (----) Jak antagonizam; (-----) Veoma jak antagonizam.

Tabela 4.10. Analiza sinergizama za kombinaciju novih HDAC inhibitora i fingolimod.

	MIA PaCa-2 ćelijska linija				Panc-1 ćelijska linija			
	Koncentracija jedinjenja ^a				Koncentracija jedinjenja ^a			
	I	II	III	IV	I	II	III	IV
HDACi (6b)	71,49 ± 11,91	41,62 ± 11,91	23,06 ± 9,75	16,27 ± 5,34	63,68 ± 4,70	25,29 ± 6,71	23,93 ± 6,23	10,30 ± 1,96
Fingolimod	79,98 ± 10,36	48,31 ± 12,04	20,57 ± 4,85	20,33 ± 5,37	79,18 ± 1,45	40,09 ± 4,97	24,56 ± 1,24	18,99 ± 0,95
6b + Fingolimod	89,28 ± 4,05	60,81 ± 20,19	34,11 ± 14,58	26,96 ± 7,59	84,96 ± 2,05	67,22 ± 5,98	37,87 ± 5,71	21,98 ± 0,78
Kombinatorijski indeks (CI)	0,780	1,422	1,648	1,070	0,847	0,945	1,232	1,141
Interakcija 6b + Fingolimod ^b	++	--	---	±	++	±	--	-
HDACi (8b)	81,22 ± 0,96	68,72 ± 2,69	36,08 ± 3,29	16,72 ± 4,28	58,10 ± 0,24	23,86 ± 9,24	4,77 ± 2,69	4,68 ± 2,33
Fingolimod	87,22 ± 2,01	52,02 ± 2,47	23,73 ± 3,92	22,83 ± 1,89	68,24 ± 4,62	41,81 ± 5,37	19,44 ± 10,97	12,27 ± 8,75
8b + Fingolimod	91,31 ± 10,08	65,06 ± 1,92	38,98 ± 5,26	26,55 ± 2,94	74,74 ± 5,61	30,91 ± 6,84	16,60 ± 7,22	9,42 ± 5,57
Kombinatorijski indeks (CI)	1,188	1,836	1,852	1,338	1,267	1,806	2,027	1,592
Interakcija 8b + Fingolimod ^b	---	---	---	--	--	---	---	---
HDACi (9b)	83,61 ± 0,27	69,42 ± 11,34	53,31 ± 17,80	44,84 ± 11,21	61,28 ± 6,91	50,93 ± 3,83	40,71 ± 3,74	24,03 ± 5,73
Fingolimod	76,58 ± 4,48	63,82 ± 6,81	43,78 ± 12,65	35,34 ± 6,99	62,10 ± 4,94	50,04 ± 3,20	31,21 ± 11,85	21,17 ± 3,33
9b + Fingolimod	89,11 ± 1,57	73,59 ± 12,40	55,57 ± 18,16	47,22 ± 9,51	76,50 ± 5,14	66,12 ± 10,76	53,41 ± 18,12	33,34 ± 10,07
Kombinatorijski indeks (CI)	0,870	1,449	1,773	1,293	0,911	0,848	0,810	1,120
Interakcija 9b + Fingolimod ^b	+	--	---	--	±	++	++	-

^a Koncentracije su prikazane u silaznom nizu (I je najviša. IV je najniža). ^b Tumačenje: (+++++) Veoma jak sinergizam; (++++) Sinergizam; (++) Umeren sinergizam; (+) Slab sinergizam; (±) Približno aditivni; (-) Slab antagonizam; (--) Umeren antagonizam; (---) Antagonizam; (----) Jak antagonizam; (-----) Veoma jak antagonizam.

Važno je napomenuti da HDAC inhibitori 6b, 8b i 9b nisu bili deo početnog skupa podataka o hemijskoj osetljivosti koji je korišćen u bioinformatičkoj analizi. Stoga, eksperimentalno detektovani sinergizmi za 6b, 8b i 9b ukazuju da su sinergizmi predviđeni bioinformatičkom analizom uglavnom rezultat direktnog inhibiranja HDAC-a, a ne sporednih efekata specifičnih HDAC inhibitora koji su uključeni u skup podataka o hemijskoj osetljivosti.

5. Zaključak

U okviru doktorske disertacije „Simulacije molekulske dinamike, dizajn, sinteza i *in vitro* ispitivanje inhibitora epigenetskih proteina kao potencijalnih antineoplastika“ izvršena su opsežna ispitivanja konformacione dinamike SIRT2 i WDR5 proteina čime su rasvetljeni kompleksni mehanizmi protein-ligand interakcija i postavljeni i validirani protokoli racionalnog dizajna novih inhibitora navedenih proteina zasnovani na poznavanju 3D-strukture ciljnog mesta. Dodatno, kroz heminformatičke i bioinformatičke studije koje se oslanjaju na rad sa velikim setovima podataka i savremene algoritme mašinskog učenja, postavljeni su i eksperimentalno validirani protokoli za razvoj novih SIRT2 inhibitora kao i identifikaciju novih ciljnih mesta dejstva u razvoju dualnih inhibitora histon-deacetilaza. Kao najznačajniji rezultati ove disertacije ističu se:

a. *Razvoj univerzalnog protokola za poboljšano uzorkovanje konformacionih stanja vezivnog mesta*

Istraživanja sprovedena u okviru ove disertacije ukazala su na višestruke načine vezivanja liganda i postojanje alternativnih konformacionih stanja vezivnog mesta SIRT2. Dok razmatranje dinamike vezivnog mesta predstavlja važan aspekt racionalnog dizajna lekova, eksperimentalno određivanje dinamike ostaje izazovan poduhvat. U okviru ove disertacije razvijen je *in silico* protokol za simulaciju dinamike vezivnog mesta. U okviru razvijenog protokola, poboljšano uzorkovanje dinamike vezivnog mesta je postignuto kroz korišćenje kolektivnih varijabli izvedenih direktno iz klasičnih MD simulacija korišćenjem tICA metode. Spori modovi dinamike vezivnog mesta su dalje ubrzani primenom dobro temperirane metadinamike uz upotrebu navedenih kolektivnih varijabli. Studijama validacije kroz primenu retrospektivnog skrininga SIRT2 inhibitora, utvrđena je sposobnost razvijenog protokola da efikasno uzorkuje alternativna konformaciona stanja nedostupna klasičnim MD simulacijama. Treba istaći da je razvijen protokol univerzalnog karaktera i da se kao takav može primeniti na različita ciljna mesta;

b. *Otkriće novog skrivenog džepa u strukturi vezivnog mesta SIRT2 i proširenje hemijskog prostora SIRT2 inhibitora kroz implementaciju alternativnog konformacionog stanja u SBVS studiju*

U okviru ispitivanja alternativnih konformacionih stanja SIRT2 primenom razvijenog protokola poboljšanog uzorkovanja otkriveno je postojanje skrivenog vezivnog džepa u nastavku džepa selektivnosti SIRT2 inhibitora. Implementacijom alternativnog konformacionog stanja SIRT2 sa otvorenim skrivenim džepom u retrospektivnu SBVS studiju, demonstriran je veliki značaj novootkrivenog džepa u objašnjavanju aktivnosti postojećih inhibitora SIRT2. Daljim uključivanjem alternativnih konformacionih stanja u prospektivnu SBVS studiju otkriven je i veliki potencijal proširenja poznatog hemijskog prostora SIRT2 inhibitora;

c. *Otkriće novog SIRT2 inhibitora sa antitumorskim potencijalom u neistraženim delovima hemijskog prostora SIRT2 inhibitora*

Eksperimentalna validacija SBVS modela sa implementiranim alternativnim konformacionim stanjima SIRT2 je izvedena testiranjem potentnosti i selektivnosti 9 virtuelnih *hit* jedinjenja preuzetih iz neistraženih delova hemijskog prostora. Sva jedinjenja su pokazala određeni procenat inhibicije SIRT2 i selektivnost u odnosu na testirane SIRT1/6 izoforme. Najpotentnija jedinjenje ND18 je pokazalo IC_{50} (SIRT2) = 58,7 μ M što je u rangu potentnosti poznatog sirtuinskog inhibitora – kaminola. Dodatno, ispitivanjem antitumorskog potencijala izdvojenog jedinjenja na MDA-MB-231 ćelijskoj liniji uočen je citostatski potencijal jedinjenja kroz zaustavljanje ćelijskog ciklusa u G2/M fazi kao i značajna reverzija morfologije MDA-MB-231 ćelija iz mezenhimalne u epitelijalnu što je bilo ispraćeno povećanjem ekspresije E-kadherina. Uzimajući u obzir značaj E-kadherina u procesu epitelijalno-mezenhimalne tranzicije, postavljena je hipoteza da jedinjenje ND18

ima sposobnost reverzije procesa čime je stvorena osnova za dalja ispitivanja antimetastatskog potencijala;

d. *Razvoj in silico platforme za dizajn novih SIRT2 inhibitora bazirane na poznavanju strukture liganda*

Rastući broj pretkliničkih istraživanja ukazuje na potencijal inhibitora SIRT2 kao novih potencijalnih terapeutika u lečenju širokog spektra oboljenja povezanih sa starenjem. Uprkos rastućem interesovanju za razvoj inhibitora SIRT2 u protekloj deceniji, nijedan od inhibitora još uvek nije ušao u klinička ispitivanja. U cilju olakšanja, ubrzanja i smanjenja troškova budućeg razvoja selektivnih i neselektivnih SIRT2 inhibitora, kreirana je *in silico* platforma za LBVS i racionalizaciju odnosa strukture i aktivnosti SIRT2 inhibitora. Platforma je uspostavljena pomoću 4 visoko validirana modela mašinskog učenja izgrađena implementacijom svih dostupnih informacija o strukturi i aktivnosti/selektivnosti poznatih SIRT2 inhibitora. Nakon temeljne validacije eksternog prediktivnog potencijala modela, izdvojena su četiri modela: binarni RF:ECFP4 model i regresioni XGBoost:ECFP4 model za predviđanje potentnosti inhibitora, kao i RF:ECFP4 model za selektivnost SIRT1/2 i DNN: deskriptorski model za predviđanje selektivnosti SIRT2/3. U cilju pružanja najbolju prakse u primeni razvijenih modela, definisani okvir za predviđanje aktivnosti/selektivnosti novih jedinjenja implementiran je u Python aplikaciju nazvanu *SIRT2i Predictor*. *SIRT2i Predictor* je opremljen intuitivnim grafičkim korisničkim interfejsom izgrađenim u web okruženju čime je omogućen rad širokoj zajednici korisnika. Sa automatskom obradom ulaznog formata (SMILES) i demonstriranom sposobnošću brzog i efikasnog evaluiranja velikih baza jedinjenja u pogledu inhibicije SIRT2 i selektivnosti SIRT1-3, *SIRT2i Predictor* nalazi veliku primenu u pružanju podrške skupocenim kampanjama visoko propusnih *in vitro* skrininga. Mogućnosti vizualizacije, koje daju mogućnost pregleda delova molekula koji pozitivno i negativno doprinose aktivnosti čine *SIRT2i Predictor* vrednim alatom za studije optimizacije vodećih jedinjenja. Sprovedena studija poređenja uz eksperimentalnu validaciju ukazala je na komplementarnost *SIRT2i Predictor*-a sa SBVS pristupom. Treba istaći univerzalnost koda generisanog za *SIRT2i Predictor* koji omogućava proširenje na seriju ciljnih mesta od interesa. Kod je deponovan u javnom repozitorijumu kom je moguće pristupiti putem linka: https://github.com/echonemanja/SIRT2i_Predictor.

e. *Sinteza novih SIRT2 inhibitora*

Na osnovu *lead* strukture ND18 identifikovane u prethodno nepoznatom delu hemijskog prostora SIRT2 inhibitora, a uz interaktivnu primenu *in silico* modela generisanih i validiranih u okviru disertacije, sintetisano je 12 novih jedinjenja, derivata fenil-proapanske, cimetine, ili fenil-sirćetne kiseline. Skoro sva jedinjenja su pokazala određeni inhibitoryni efekat na SIRT2 i selektivnost u odnosu na testiranu SIRT1. Polovina sintetisanih jedinjenja je pokazala procenat inhibicije SIRT2 veći od 50% pri testiranoj koncentraciji od 100 μ M što ih čini obećavajućim kandidatima za dalja *in vitro* ispitivanja antitumorskog potencijala. Među sintetisanim jedinjenjima se posebno ističe klasa aril-1,2,3-triazola kao potpuno nove hemijske grupe SIRT2 inhibitora;

f. *Otkriće novog puta alosterne komunikacije unutar WDR5 i razvoj protokola za dizajn optofarmakoloških WDR5 inhibitora*

Istraživanjima sprovedenim na grupi potpuno nove klase fotoreaktivnih peptidomimetskih inhibitora protein-protein interakcije WDR5, okarakterisano je konformaciono ponašanje WDR5 u prisustvu različitih liganada i postavljen je protokol dizajna novih predstavnika ove grupe optofarmakoloških liganada zasnovan na MD simulacijama i proračunima slobodne energije vezivanja MM/PBSA metodom. Primenom metadinamičkih simulacija,

kvantifikovane su energetske promena procesa fotoizomerizacije azobenzena u okviru ispitivanog cikličnog peptidomimetika čime je pružen atomistički uvid u mehanizme procesa *in situ* izomerizacije što je omogućilo kompletiranje nepotpune slike selektivnosti *cis* i *trans* izomera dobijene eksperimentalnim metodama. Na osnovu analize dinamičke mreže interakcija aminokiselinskih rezidua izvedene iz opsežnih MD simulacija uspostavljen je model alosterne komunikacije dva vezivana mesta (WIN i WBM) kroz centralni kanal WDR5. Novootkrivenom alosternom komunikacijom postavljena je osnova za dizajn potentnijih WIN inhibitora WDR5 kroz efikasnije ometanje komunikacije između dva vezivna mesta i posledične disocijacije proteina multiproteinskih kompleksa koji interaguju sa WBM mestom WDR5 proteina;

g. *Otkriće novih ciljnih mesta za razvoj dualnih inhibitora histon-deacetilaza u terapiji karcinoma pankreasa*

U eri personalizovane medicine, sistemski prilagođena kombinovana terapija ima veliki potencijal unapređenja nedovoljno efikasne hemioterapije adenokarcinoma pankreasa. Usled postojanja velikog broja mogućih kombinacija različitih bioaktivnih jedinjenja, eksperimentalna identifikacija sinergizma predstavlja teško dostižan cilj. U okviru istraživanja ovog dela disertacije razvijen je bioinformatički protokol za predviđanje sinergističkih kombinacija HDAC inhibitora (uključujući SIRT inhibitore) i ostalih bioaktivnih jedinjenja u tretmanu adenokarcinoma pankreasa. Opsežnom bioinformatičkom analizom podataka o osetljivosti na bioaktivna jedinjenja, zajedno sa podacima o bazalnoj ekspresiji gena u ćelijskim linijama poreklom iz adenokarcinoma pankreasa, identifikovan je sfingolipidni signalni puta zajedno sa povezanim efektorskim putevima kao novo ciljnih mesta za budući razvoj dualnih HDAC inhibitora ili razvoj kombinovane terapije. Eksperimentalna validacija protokola dovela je do detekcije sinergizma između HDAC inhibitora i ROCK inhibitora, kao i sinergizma između HDAC inhibitora i fingolimoda (funkcionalni antagonist podtipa 1 receptora za S1P).

6. Literatura

- (1) Mulholland, A. J. Introduction. Biomolecular Simulation. *J R Soc Interface* **2008**, *5* (Suppl 3), S169–S172. <https://doi.org/10.1098/rsif.2008.0385.focus>.
- (2) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biology* **2011**, *9* (1), 71. <https://doi.org/10.1186/1741-7007-9-71>.
- (3) Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* **2016**, *17* (2). <https://doi.org/10.3390/ijms17020144>.
- (4) Janin, J. Protein-Protein Recognition. *Progress in Biophysics and Molecular Biology* **1995**, *64* (2), 145–166. [https://doi.org/10.1016/S0079-6107\(96\)00001-6](https://doi.org/10.1016/S0079-6107(96)00001-6).
- (5) Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **2019**, *20* (11). <https://doi.org/10.3390/ijms20112783>.
- (6) Lu, H.; Tonge, P. J. Drug–Target Residence Time: Critical Information for Lead Optimization. *Current Opinion in Chemical Biology* **2010**, *14* (4), 467–474. <https://doi.org/10.1016/j.cbpa.2010.06.176>.
- (7) Liu, X.; Golden, L. C.; Lopez, J. A.; Shepherd, T. R.; Yu, L.; Fuentes, E. J. Conformational Dynamics and Cooperativity Drive the Specificity of a Protein-Ligand Interaction. *Biophysical Journal* **2019**, *116* (12), 2314–2330. <https://doi.org/10.1016/j.bpj.2019.05.008>.
- (8) Petrović, D.; Risso, V. A.; Kamerlin, S. C. L.; Sanchez-Ruiz, J. M. Conformational Dynamics and Enzyme Evolution. *Journal of The Royal Society Interface* **2018**, *15* (144), 20180330. <https://doi.org/10.1098/rsif.2018.0330>.
- (9) Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem. Rev.* **2016**, *116* (11), 6516–6551. <https://doi.org/10.1021/acs.chemrev.5b00562>.
- (10) Dror, R. O.; Dirks, R. M.; Grossman, J. p.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41* (1), 429–452. <https://doi.org/10.1146/annurev-biophys-042910-155245>.
- (11) Fenwick, R. B.; van den Bedem, H.; Fraser, J. S.; Wright, P. E. Integrated Description of Protein Dynamics from Room-Temperature X-Ray Crystallography and NMR. *Proc Natl Acad Sci U S A* **2014**, *111* (4), E445–454. <https://doi.org/10.1073/pnas.1323440111>.
- (12) Qin, H.; Lim, L.; Song, J. Protein Dynamics at Eph Receptor-Ligand Interfaces as Revealed by Crystallography, NMR and MD Simulations. *BMC Biophys* **2012**, *5*, 2. <https://doi.org/10.1186/2046-1682-5-2>.
- (13) Aci-Sèche, S.; Ziada, S.; Braka, A.; Arora, R.; Bonnet, P. Advanced Molecular Dynamics Simulation Methods for Kinase Drug Discovery. *Future Med Chem* **2016**, *8* (5), 545–566. <https://doi.org/10.4155/fmc.16.9>.
- (14) Borhani, D. W.; Shaw, D. E. The Future of Molecular Dynamics Simulations in Drug Discovery. *J Comput Aided Mol Des* **2012**, *26* (1), 15–26. <https://doi.org/10.1007/s10822-011-9517-y>.
- (15) Andrew Leach. *Molecular Modelling: Principles and Applications - 2nd Edition*.
- (16) Sanjeev, A.; Kaushik, S.; Meher, B. R.; Mattaparthi, V. S. K. Disorder in Proteins. In *Reference Module in Life Sciences*; Elsevier, 2018. <https://doi.org/10.1016/B978-0-12-809633-8.20271-9>.
- (17) Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98–103. <https://doi.org/10.1103/PhysRev.159.98>.
- (18) Saman Alavi. *Molecular Simulations: Fundamentals and Practice* | Wiley.
- (19) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High-Resolution Computer Models of a Plasma. *Journal of Computational Physics* **1974**, *14* (2), 148–158. [https://doi.org/10.1016/0021-9991\(74\)90010-2](https://doi.org/10.1016/0021-9991(74)90010-2).

- (20) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *The Journal of Chemical Physics* **1982**, *76* (1), 637–649. <https://doi.org/10.1063/1.442716>.
- (21) Salo-Ahen, O. M. H.; Alanko, I.; Bhadane, R.; Bonvin, A. M. J. J.; Honorato, R. V.; Hossain, S.; Juffer, A. H.; Kabedev, A.; Lahtela-Kakkonen, M.; Larsen, A. S.; Lescrinier, E.; Marimuthu, P.; Mirza, M. U.; Mustafa, G.; Nunes-Alves, A.; Pansar, T.; Saadabadi, A.; Singaravelu, K.; Vanmeert, M. Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. *Processes* **2021**, *9* (1), 71. <https://doi.org/10.3390/pr9010071>.
- (22) *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Castleman, A. W., Toennies, J. P., Yamanouchi, K., Zinth, W., Series Eds.; Springer Series in CHEMICAL PHYSICS; Springer: Berlin, Heidelberg, 2007; Vol. 86. <https://doi.org/10.1007/978-3-540-38448-9>.
- (23) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hémin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 044130. <https://doi.org/10.1063/5.0014475>.
- (24) Nam, K. H. Molecular Dynamics—From Small Molecules to Macromolecules. *International Journal of Molecular Sciences* **2021**, *22* (7), 3761. <https://doi.org/10.3390/ijms22073761>.
- (25) Pang, Y. T.; Pavlova, A.; Tajkhorshid, E.; Gumbart, J. C. Parameterization of a Drug Molecule with a Halogen σ -Hole Particle Using FfTK: Implementation, Testing, and Comparison. *J. Chem. Phys.* **2020**, *153* (16), 164104. <https://doi.org/10.1063/5.0022802>.
- (26) Groenhof, G. Introduction to QM/MM Simulations. In *Biomolecular Simulations: Methods and Protocols*; Monticelli, L., Salonen, E., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2013; pp 43–66. https://doi.org/10.1007/978-1-62703-017-5_3.
- (27) Chen, J.; Kato, J.; Harper, J. B.; Shao, Y.; Ho, J. On the Accuracy of QM/MM Models: A Systematic Study of Intramolecular Proton Transfer Reactions of Amino Acids in Water. *J Phys Chem B* **2021**, *125* (32), 9304–9316. <https://doi.org/10.1021/acs.jpcc.1c04876>.
- (28) Sun, S.; Huggins, D. J. Assessing the Effect of Forcefield Parameter Sets on the Accuracy of Relative Binding Free Energy Calculations. *Front Mol Biosci* **2022**, *9*, 972162. <https://doi.org/10.3389/fmolb.2022.972162>.
- (29) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D. Current Status of Protein Force Fields for Molecular Dynamics. *Methods Mol Biol* **2015**, *1215*, 47–71. https://doi.org/10.1007/978-1-4939-1465-4_3.
- (30) Pavlova, A.; Gumbart, J. C. Parametrization of Macrolide Antibiotics Using the Force Field Toolkit. *J Comput Chem* **2015**, *36* (27), 2052–2063. <https://doi.org/10.1002/jcc.24043>.
- (31) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J Comput Chem* **2013**, *34* (32), 2757–2770. <https://doi.org/10.1002/jcc.23422>.
- (32) Allison, J. R. Computational Methods for Exploring Protein Conformations. *Biochem Soc Trans* **2020**, *48* (4), 1707–1724. <https://doi.org/10.1042/BST20200193>.
- (33) Pampel, B.; Valsson, O. Improving the Efficiency of Variationally Enhanced Sampling with Wavelet-Based Bias Potentials. *J. Chem. Theory Comput.* **2022**, *18* (7), 4127–4141. <https://doi.org/10.1021/acs.jctc.2c00197>.
- (34) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced Sampling in Molecular Dynamics. *The Journal of Chemical Physics* **2019**, *151* (7), 070902. <https://doi.org/10.1063/1.5109531>.
- (35) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nature Reviews Physics* **2020**, *2* (4), 200–212. <https://doi.org/10.1038/s42254-020-0153-0>.

- (36) Damjanovic, J.; Miao, J.; Huang, H.; Lin, Y.-S. Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations. *Chem Rev* **2021**, *121* (4), 2292–2324. <https://doi.org/10.1021/acs.chemrev.0c01087>.
- (37) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- (38) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. <https://doi.org/10.1103/PhysRevLett.100.020603>.
- (39) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Computational Molecular Science* **2011**, *1* (5), 826–843. <https://doi.org/10.1002/wcms.31>.
- (40) Barducci, A.; Bonomi, M.; Parrinello, M. Linking Well-Tempered Metadynamics Simulations with Experiments. *Biophysical Journal* **2010**, *98* (9), L44–L46. <https://doi.org/10.1016/j.bpj.2010.01.033>.
- (41) Sutto, L.; Marsili, S.; Gervasio, F. L. New Advances in Metadynamics. *WIREs Computational Molecular Science* **2012**, *2* (5), 771–779. <https://doi.org/10.1002/wcms.1103>.
- (42) Willems, H.; De Cesco, S.; Svensson, F. Computational Chemistry on a Budget: Supporting Drug Discovery with Limited Resources. *J Med Chem* **2020**, *63* (18), 10158–10169. <https://doi.org/10.1021/acs.jmedchem.9b02126>.
- (43) Morrison, C. Fresh from the Biotech Pipeline—2018. *Nature Biotechnology* **2019**, *37* (2), 118–123. <https://doi.org/10.1038/s41587-019-0021-6>.
- (44) Carracedo-Reboredo, P.; Liñares-Blanco, J.; Rodríguez-Fernández, N.; Cedrón, F.; Novoa, F. J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C. A Review on Machine Learning Approaches and Trends in Drug Discovery. *Computational and Structural Biotechnology Journal* **2021**, *19*, 4538–4558. <https://doi.org/10.1016/j.csbj.2021.08.011>.
- (45) Ou-Yang, S.; Lu, J.; Kong, X.; Liang, Z.; Luo, C.; Jiang, H. Computational Drug Discovery. *Acta Pharmacol Sin* **2012**, *33* (9), 1131–1140. <https://doi.org/10.1038/aps.2012.109>.
- (46) David, A.; Islam, S.; Tankhilevich, E.; Sternberg, M. J. E. The AlphaFold Database of Protein Structures: A Biologist’s Guide. *Journal of Molecular Biology* **2022**, *434* (2), 167336. <https://doi.org/10.1016/j.jmb.2021.167336>.
- (47) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat Rev Drug Discov* **2018**, *17* (8), 547–558. <https://doi.org/10.1038/nrd.2018.93>.
- (48) Dunn, M. F. Protein–Ligand Interactions: General Description. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd, 2010. <https://doi.org/10.1002/9780470015902.a0001340.pub2>.
- (49) Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. *Annual Review of Physical Chemistry* **2019**, *70* (1), 143–171. <https://doi.org/10.1146/annurev-physchem-042018-052340>.
- (50) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat Commun* **2017**, *8* (1), 2276. <https://doi.org/10.1038/s41467-017-02258-w>.
- (51) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends in Biochemical Sciences* **2010**, *35* (10), 539–546. <https://doi.org/10.1016/j.tibs.2010.04.009>.
- (52) Limongelli, V. Ligand Binding Free Energy and Kinetics Calculation in 2020. *WIREs Computational Molecular Science* **2020**, *10* (4), e1455. <https://doi.org/10.1002/wcms.1455>.
- (53) Huang, S.-Y.; Zou, X. Advances and Challenges in Protein-Ligand Docking. *Int J Mol Sci* **2010**, *11* (8), 3016–3034. <https://doi.org/10.3390/ijms11083016>.

- (54) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins* **2002**, *47* (4), 409–443. <https://doi.org/10.1002/prot.10115>.
- (55) Knegtel, R. M. A.; Antoon, J.; Rullmann, C.; Boelens, R.; Kaptein, R. MONTY: A Monte Carlo Approach to Protein-DNA Recognition. *Journal of Molecular Biology* **1994**, *235* (1), 318–324. [https://doi.org/10.1016/S0022-2836\(05\)80035-X](https://doi.org/10.1016/S0022-2836(05)80035-X).
- (56) Gardiner, E. J.; Willett, P.; Artymiuk, P. J. Protein Docking Using a Genetic Algorithm. *Proteins* **2001**, *44* (1), 44–56. <https://doi.org/10.1002/prot.1070>.
- (57) Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdiscip Sci Comput Life Sci* **2019**, *11* (2), 320–328. <https://doi.org/10.1007/s12539-019-00327-w>.
- (58) Gioia, D.; Bertazzo, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules* **2017**, *22* (11), 2029. <https://doi.org/10.3390/molecules22112029>.
- (59) Guterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J Chem Inf Model* **2020**, *60* (4), 2189–2198. <https://doi.org/10.1021/acs.jcim.0c00057>.
- (60) Maia, E. H. B.; Assis, L. C.; de Oliveira, T. A.; da Silva, A. M.; Taranto, A. G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in Chemistry* **2020**, *8*.
- (61) Wang, Z.; Sun, H.; Shen, C.; Hu, X.; Gao, J.; Li, D.; Cao, D.; Hou, T. Combined Strategies in Structure-Based Virtual Screening. *Phys. Chem. Chem. Phys.* **2020**, *22* (6), 3149–3159. <https://doi.org/10.1039/C9CP06303J>.
- (62) Varela-Rial, A.; Majewski, M.; De Fabritiis, G. Structure Based Virtual Screening: Fast and Slow. *WIREs Computational Molecular Science* **2022**, *12* (2), e1544. <https://doi.org/10.1002/wcms.1544>.
- (63) Roy, K.; Mitra, I. On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design. *Comb Chem High Throughput Screen* **2011**, *14* (6), 450–474. <https://doi.org/10.2174/138620711795767893>.
- (64) Truchon, J.-F.; Bayly, C. I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem. *J. Chem. Inf. Model.* **2007**, *47* (2), 488–508. <https://doi.org/10.1021/ci600426e>.
- (65) Gimeno, A.; Ojeda-Montes, M. J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int J Mol Sci* **2019**, *20* (6), 1375. <https://doi.org/10.3390/ijms20061375>.
- (66) Kralj, S.; Jukič, M.; Bren, U. Molecular Filters in Medicinal Chemistry. *Encyclopedia* **2023**, *3* (2), 501–511. <https://doi.org/10.3390/encyclopedia3020035>.
- (67) Li, D.; Ji, B. Protein Conformational Transitions Coupling with Ligand Interactions: Simulations from Molecules to Medicine. *Medicine in Novel Technology and Devices* **2019**, *3*, 100026. <https://doi.org/10.1016/j.medntd.2019.100026>.
- (68) Spyraakis, F.; Benedetti, P.; Decherchi, S.; Rocchia, W.; Cavalli, A.; Alcaro, S.; Ortuso, F.; Baroni, M.; Cruciani, G. A Pipeline To Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and Fingerprints for Ligand and Proteins. *J. Chem. Inf. Model.* **2015**, *55* (10), 2256–2274. <https://doi.org/10.1021/acs.jcim.5b00169>.
- (69) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52* (5), 1262–1274. <https://doi.org/10.1021/ci2005934>.
- (70) Craig, I. R.; Essex, J. W.; Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* **2010**, *50* (4), 511–524. <https://doi.org/10.1021/ci900407c>.

- (71) Jubb, H.; Blundell, T. L.; Ascher, D. B. Flexibility and Small Pockets at Protein–Protein Interfaces: New Insights into Druggability. *Prog Biophys Mol Biol* **2015**, *119* (1), 2–9. <https://doi.org/10.1016/j.pbiomolbio.2015.01.009>.
- (72) Rakers, C.; Bermudez, M.; Keller, B. G.; Mortier, J.; Wolber, G. Computational Close up on Protein–Protein Interactions: How to Unravel the Invisible Using Molecular Dynamics Simulations? *WIREs Computational Molecular Science* **2015**, *5* (5), 345–359. <https://doi.org/10.1002/wcms.1222>.
- (73) Perricone, U.; Gulotta, M. R.; Lombino, J.; Parrino, B.; Cascioferro, S.; Diana, P.; Cirrincione, G.; Padova, A. An Overview of Recent Molecular Dynamics Applications as Medicinal Chemistry Tools for the Undruggable Site Challenge. *Med. Chem. Commun.* **2018**, *9* (6), 920–936. <https://doi.org/10.1039/C8MD00166A>.
- (74) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev* **2019**, *119* (16), 9478–9508. <https://doi.org/10.1021/acs.chemrev.9b00055>.
- (75) Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H. Calculation of Binding Free Energies. *Methods Mol Biol* **2015**, *1215*, 173–209. https://doi.org/10.1007/978-1-4939-1465-4_9.
- (76) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57* (12), 2911–2937. <https://doi.org/10.1021/acs.jcim.7b00564>.
- (77) Miller, B. R. I.; McGee, T. D. Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314–3321. <https://doi.org/10.1021/ct300418h>.
- (78) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol Rev* **2014**, *66* (1), 334–395. <https://doi.org/10.1124/pr.112.007336>.
- (79) Qian, T.; Zhu, S.; Hoshida, Y. Use of Big Data in Drug Development for Precision Medicine: An Update. *Expert Rev Precis Med Drug Dev* **2019**, *4* (3), 189–200. <https://doi.org/10.1080/23808993.2019.1617632>.
- (80) Zhu, H. Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annual Review of Pharmacology and Toxicology* **2020**, *60* (1), 573–589. <https://doi.org/10.1146/annurev-pharmtox-010919-023324>.
- (81) Dixon, S. L.; Duan, J.; Smith, E.; Von Bargen, C. D.; Sherman, W.; Repasky, M. P. AutoQSAR: An Automated Machine Learning Tool for Best-Practice Quantitative Structure–Activity Relationship Modeling. *Future Med Chem* **2016**, *8* (15), 1825–1839. <https://doi.org/10.4155/fmc-2016-0093>.
- (82) Sabe, V. T.; Ntombela, T.; Jhamba, L. A.; Maguire, G. E. M.; Govender, T.; Naicker, T.; Kruger, H. G. Current Trends in Computer Aided Drug Design and a Highlight of Drugs Discovered via Computational Techniques: A Review. *European Journal of Medicinal Chemistry* **2021**, *224*, 113705. <https://doi.org/10.1016/j.ejmech.2021.113705>.
- (83) Winkler, D. A. Sparse QSAR Modelling Methods for Therapeutic and Regenerative Medicine. *J Comput Aided Mol Des* **2018**, *32* (4), 497–509. <https://doi.org/10.1007/s10822-018-0106-1>.
- (84) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57* (12), 4977–5010. <https://doi.org/10.1021/jm4004285>.
- (85) N. Muratov, E.; Bajorath, J.; P. Sheridan, R.; V. Tetko, I.; Filimonov, D.; Poroikov, V.; I. Oprea, T.; I. Baskin, I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; A. Winkler, D.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A.

- QSAR without Borders. *Chemical Society Reviews* **2020**, *49* (11), 3525–3564. <https://doi.org/10.1039/D0CS00098A>.
- (86) Bertoni, M.; Duran-Frigola, M.; Badia-i-Mompel, P.; Pauls, E.; Orozco-Ruiz, M.; Guitart-Pla, O.; Alcalde, V.; Diaz, V. M.; Berenguer-Llargo, A.; Brun-Heath, I.; Villegas, N.; de Herreros, A. G.; Aloy, P. Bioactivity Descriptors for Uncharacterized Chemical Compounds. *Nat Commun* **2021**, *12* (1), 3932. <https://doi.org/10.1038/s41467-021-24150-4>.
- (87) Dong, J.; Cao, D.-S.; Miao, H.-Y.; Liu, S.; Deng, B.-C.; Yun, Y.-H.; Wang, N.-N.; Lu, A.-P.; Zeng, W.-B.; Chen, A. F. ChemDes: An Integrated Web-Based Platform for Molecular Descriptor and Fingerprint Computation. *Journal of Cheminformatics* **2015**, *7* (1), 60. <https://doi.org/10.1186/s13321-015-0109-z>.
- (88) Andrade, C. H.; Pasqualoto, K. F. M.; Ferreira, E. I.; Hopfinger, A. J. 4D-QSAR: Perspectives in Drug Design. *Molecules* **2010**, *15* (5), 3281–3294. <https://doi.org/10.3390/molecules15053281>.
- (89) Gao, K.; Nguyen, D. D.; Sresht, V.; Mathiowetz, A. M.; Tu, M.; Wei, G.-W. Are 2D Fingerprints Still Valuable for Drug Discovery? *Phys. Chem. Chem. Phys.* **2020**, *22* (16), 8373–8390. <https://doi.org/10.1039/D0CP00305K>.
- (90) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J Chem Inf Comput Sci* **2002**, *42* (6), 1273–1280. <https://doi.org/10.1021/ci010132r>.
- (91) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J Chem Inf Model* **2010**, *50* (5), 742–754. <https://doi.org/10.1021/ci100050t>.
- (92) Gramatica, P. Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *IJQSPR* **2020**, *5* (3), 61–97. <https://doi.org/10.4018/IJQSPR.20200701.oa1>.
- (93) Tsou, L. K.; Yeh, S.-H.; Ueng, S.-H.; Chang, C.-P.; Song, J.-S.; Wu, M.-H.; Chang, H.-F.; Chen, S.-R.; Shih, C.; Chen, C.-T.; Ke, Y.-Y. Comparative Study between Deep Learning and QSAR Classifications for TNBC Inhibitors and Novel GPCR Agonist Discovery. *Sci Rep* **2020**, *10* (1), 16771. <https://doi.org/10.1038/s41598-020-73681-1>.
- (94) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics* **2010**, *29* (6–7), 476–488. <https://doi.org/10.1002/minf.201000061>.
- (95) *ChEMBL Database*. <https://www.ebi.ac.uk/chembl/> (accessed 2021-09-03).
- (96) *Validation of (Q)SAR Models - OECD*. <https://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.htm> (accessed 2022-08-20).
- (97) Bosc, N.; Atkinson, F.; Felix, E.; Gaulton, A.; Hersey, A.; Leach, A. R. Large Scale Comparison of QSAR and Conformal Prediction Methods and Their Applications in Drug Discovery. *Journal of Cheminformatics* **2019**, *11* (1), 4. <https://doi.org/10.1186/s13321-018-0325-4>.
- (98) Gramatica, P. On the Development and Validation of QSAR Models. *Methods Mol Biol* **2013**, *930*, 499–526. https://doi.org/10.1007/978-1-62703-059-5_21.
- (99) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **2009**, *49* (7), 1669–1678. <https://doi.org/10.1021/ci900115y>.
- (100) Schüürmann, G.; Ebert, R.-U.; Chen, J.; Wang, B.; Kühne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient Test Set Activity Mean vs Training Set Activity Mean. *J Chem Inf Model* **2008**, *48* (11), 2140–2145. <https://doi.org/10.1021/ci800253u>.
- (101) Ojha, P. K.; Mitra, I.; Das, R. N.; Roy, K. Further Exploring Rm2 Metrics for Validation of QSPR Models. *Chemometrics and Intelligent Laboratory Systems* **2011**, *107* (1), 194–205. <https://doi.org/10.1016/j.chemolab.2011.03.011>.
- (102) Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N. Some Case Studies on Application of “Rm2” Metrics for Judging Quality of Quantitative Structure–Activity Relationship Predictions: Emphasis on Scaling of Response Data. *Journal of Computational Chemistry* **2013**, *34* (12), 1071–1082. <https://doi.org/10.1002/jcc.23231>.

- (103) Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. *J. Chem. Inf. Model.* **2011**, *51* (9), 2320–2335. <https://doi.org/10.1021/ci200211n>.
- (104) Li, X.; Kleinstreuer, N. C.; Fourches, D. Hierarchical Quantitative Structure-Activity Relationship Modeling Approach for Integrating Binary, Multiclass, and Regression Models of Acute Oral Systemic Toxicity. *Chem Res Toxicol* **2020**, *33* (2), 353–366. <https://doi.org/10.1021/acs.chemrestox.9b00259>.
- (105) Matveieva, M.; Polishchuk, P. Benchmarks for Interpretation of QSAR Models. *Journal of Cheminformatics* **2021**, *13* (1), 41. <https://doi.org/10.1186/s13321-021-00519-x>.
- (106) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach Learn* **1995**, *20* (3), 273–297. <https://doi.org/10.1007/BF00994018>.
- (107) Rodríguez-Pérez, R.; Bajorath, J. Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. *J Comput Aided Mol Des* **2022**, *36* (5), 355–362. <https://doi.org/10.1007/s10822-022-00442-9>.
- (108) *1.4. Support Vector Machines*. scikit-learn. <https://scikit-learn/stable/modules/svm.html> (accessed 2023-06-02).
- (109) Mucherino, A.; Papajorgji, P. J.; Pardalos, P. M. K-Nearest Neighbor Classification. In *Data Mining in Agriculture*; Springer Optimization and Its Applications; Springer New York: New York, NY, 2009; Vol. 34, pp 83–106. https://doi.org/10.1007/978-0-387-88615-2_4.
- (110) Zhang, Z. Introduction to Machine Learning: K-Nearest Neighbors. *Ann Transl Med* **2016**, *4* (11), 218. <https://doi.org/10.21037/atm.2016.03.37>.
- (111) *1.6. Nearest Neighbors*. scikit-learn. <https://scikit-learn/stable/modules/neighbors.html> (accessed 2023-06-02).
- (112) Breiman, L. Random Forests. *Machine Learning* **2001**, *45* (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- (113) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships. *J Chem Inf Model* **2015**, *55* (2), 263–274. <https://doi.org/10.1021/ci500747n>.
- (114) Ayyadevara, V. K. Gradient Boosting Machine. In *Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R*; Ayyadevara, V. K., Ed.; Apress: Berkeley, CA, 2018; pp 117–134. https://doi.org/10.1007/978-1-4842-3564-5_6.
- (115) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **2001**, *29* (5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- (116) Sheridan, R. P.; Wang, W. M.; Liaw, A.; Ma, J.; Gifford, E. M. Extreme Gradient Boosting as a Method for Quantitative Structure-Activity Relationships. *J Chem Inf Model* **2016**, *56* (12), 2353–2360. <https://doi.org/10.1021/acs.jcim.6b00591>.
- (117) Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *Journal of Big Data* **2021**, *8* (1), 53. <https://doi.org/10.1186/s40537-021-00444-8>.
- (118) Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The Rise of Deep Learning in Drug Discovery. *Drug Discovery Today* **2018**, *23* (6), 1241–1250. <https://doi.org/10.1016/j.drudis.2018.01.039>.
- (119) Rodríguez-Pérez, R.; Miljković, F.; Bajorath, J. Machine Learning in Chemoinformatics and Medicinal Chemistry. *Annual Review of Biomedical Data Science* **2022**, *5* (1), 43–65. <https://doi.org/10.1146/annurev-biodatasci-122120-124216>.
- (120) Xu, Y. Deep Neural Networks (DNNs) for QSAR Quantitative Structure Activity Relationship (QSAR). In *Artificial Intelligence in Drug Design*; Heifetz, A., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2022; pp 233–260. https://doi.org/10.1007/978-1-0716-1787-8_10.

- (121) Goldberg, A. D.; Allis, C. D.; Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **2007**, *128* (4), 635–638. <https://doi.org/10.1016/j.cell.2007.02.006>.
- (122) Deans, C.; Maggert, K. A. What Do You Mean, “Epigenetic”? *Genetics* **2015**, *199* (4), 887–896. <https://doi.org/10.1534/genetics.114.173492>.
- (123) Tronick, E.; Hunter, R. G. Waddington, Dynamic Systems, and Epigenetics. *Front Behav Neurosci* **2016**, *10*, 107. <https://doi.org/10.3389/fnbeh.2016.00107>.
- (124) Potaczek, D. P.; Alashkar Alhamwe, B.; Miethe, S.; Garn, H. Epigenetic Mechanisms in Allergy Development and Prevention. *Handb Exp Pharmacol* **2022**, *268*, 331–357. https://doi.org/10.1007/164_2021_475.
- (125) Fernandez, A.; O’Leary, C.; O’Byrne, K. J.; Burgess, J.; Richard, D. J.; Suraweera, A. Epigenetic Mechanisms in DNA Double Strand Break Repair: A Clinical Review. *Frontiers in Molecular Biosciences* **2021**, *8*.
- (126) Lu, Y.; Chan, Y.-T.; Tan, H.-Y.; Li, S.; Wang, N.; Feng, Y. Epigenetic Regulation in Human Cancer: The Potential Role of Epi-Drug in Cancer Therapy. *Molecular Cancer* **2020**, *19* (1), 79. <https://doi.org/10.1186/s12943-020-01197-3>.
- (127) Kulis, M.; Esteller, M. DNA Methylation and Cancer. *Adv Genet* **2010**, *70*, 27–56. <https://doi.org/10.1016/B978-0-12-380866-0.60002-2>.
- (128) Jin, B.; Li, Y.; Robertson, K. D. DNA Methylation. *Genes Cancer* **2011**, *2* (6), 607–617. <https://doi.org/10.1177/1947601910393957>.
- (129) Rasmussen, K. D.; Helin, K. Role of TET Enzymes in DNA Methylation, Development, and Cancer. *Genes Dev* **2016**, *30* (7), 733–750. <https://doi.org/10.1101/gad.276568.115>.
- (130) Bannister, A. J.; Kouzarides, T. Regulation of Chromatin by Histone Modifications. *Cell Res* **2011**, *21* (3), 381–395. <https://doi.org/10.1038/cr.2011.22>.
- (131) Wei, J.-W.; Huang, K.; Yang, C.; Kang, C.-S. Non-Coding RNAs as Regulators in Epigenetics (Review). *Oncology Reports* **2017**, *37* (1), 3–9. <https://doi.org/10.3892/or.2016.5236>.
- (132) Huang, W.; Li, H.; Yu, Q.; Xiao, W.; Wang, D. O. LncRNA-Mediated DNA Methylation: An Emerging Mechanism in Cancer and Beyond. *Journal of Experimental & Clinical Cancer Research* **2022**, *41* (1), 100. <https://doi.org/10.1186/s13046-022-02319-z>.
- (133) Taniguchi, Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. *Int J Mol Sci* **2016**, *17* (11), 1849. <https://doi.org/10.3390/ijms17111849>.
- (134) Damiani, E.; Duran, M. N.; Mohan, N.; Rajendran, P.; Dashwood, R. H. Targeting Epigenetic ‘Readers’ with Natural Compounds for Cancer Interception. *J Cancer Prev* **2020**, *25* (4), 189–203. <https://doi.org/10.15430/JCP.2020.25.4.189>.
- (135) Lee, J. E.; Kim, M.-Y. Cancer Epigenetics: Past, Present and Future. *Semin Cancer Biol* **2022**, *83*, 4–14. <https://doi.org/10.1016/j.semcancer.2021.03.025>.
- (136) Roshdy, E.; Mustafa, M.; Shaltout, A. E.-R.; Radwan, M. O.; Ibrahim, M. A. A.; Soliman, M. E.; Fujita, M.; Otsuka, M.; Ali, T. F. S. Selective SIRT2 Inhibitors as Promising Anticancer Therapeutics: An Update from 2016 to 2020. *European Journal of Medicinal Chemistry* **2021**, *224*, 113709. <https://doi.org/10.1016/j.ejmech.2021.113709>.
- (137) Aho, E. R.; Weissmiller, A. M.; Fesik, S. W.; Tansey, W. P. Targeting WDR5: A WINning Anti-Cancer Strategy? *Epigenet Insights* **2019**, *12*. <https://doi.org/10.1177/2516865719865282>.
- (138) Guarnaccia, A. duPuy; Tansey, W. P. Moonlighting with WDR5: A Cellular Multitasker. *J Clin Med* **2018**, *7* (2). <https://doi.org/10.3390/jcm7020021>.
- (139) Chen, G.; Huang, P.; Hu, C. The Role of SIRT2 in Cancer: A Novel Therapeutic Target. *International Journal of Cancer* **2020**, *147* (12), 3297–3304. <https://doi.org/10.1002/ijc.33118>.
- (140) Siladi, A. J.; Wang, J.; Florian, A. C.; Thomas, L. R.; Creighton, J. H.; Matlock, B. K.; Flaherty, D. K.; Lorey, S. L.; Howard, G. C.; Fesik, S. W.; Weissmiller, A. M.; Liu, Q.;

- Tansey, W. P. WIN Site Inhibition Disrupts a Subset of WDR5 Function. *Sci Rep* **2022**, *12*, 1848. <https://doi.org/10.1038/s41598-022-05947-9>.
- (141) North, B. J.; Marshall, B. L.; Borra, M. T.; Denu, J. M.; Verdin, E. The Human Sir2 Ortholog, SIRT2, Is an NAD⁺-Dependent Tubulin Deacetylase. *Mol Cell* **2003**, *11* (2), 437–444. [https://doi.org/10.1016/s1097-2765\(03\)00038-8](https://doi.org/10.1016/s1097-2765(03)00038-8).
- (142) Gomes, P.; Outeiro, T. F.; Cavadas, C. Emerging Role of Sirtuin 2 in the Regulation of Mammalian Metabolism. *Trends in Pharmacological Sciences* **2015**, *36* (11), 756–768. <https://doi.org/10.1016/j.tips.2015.08.001>.
- (143) Inoue, T.; Hiratsuka, M.; Osaki, M.; Yamada, H.; Kishimoto, I.; Yamaguchi, S.; Nakano, S.; Katoh, M.; Ito, H.; Oshimura, M. SIRT2, a Tubulin Deacetylase, Acts to Block the Entry to Chromosome Condensation in Response to Mitotic Stress. *Oncogene* **2007**, *26* (7), 945–957. <https://doi.org/10.1038/sj.onc.1209857>.
- (144) Zhou, Z.; Ma, T.; Zhu, Q.; Xu, Y.; Zha, X. Recent Advances in Inhibitors of Sirtuin1/2: An Update and Perspective. *Future Medicinal Chemistry* **2018**, *10* (8), 907–934. <https://doi.org/10.4155/fmc-2017-0207>.
- (145) Kudo, N.; Ito, A.; Arata, M.; Nakata, A.; Yoshida, M. Identification of a Novel Small Molecule That Inhibits Deacetylase but Not Defatty-Acylase Reaction Catalysed by SIRT2. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2018**, *373* (1748), 20170070. <https://doi.org/10.1098/rstb.2017.0070>.
- (146) Feldman, J. L.; Dittenhafer-Reed, K. E.; Kudo, N.; Thelen, J. N.; Ito, A.; Yoshida, M.; Denu, J. M. Kinetic and Structural Basis for Acyl-Group Selectivity and NAD⁺-Dependence in Sirtuin-Catalyzed Deacylation. *Biochemistry* **2015**, *54* (19), 3037–3050. <https://doi.org/10.1021/acs.biochem.5b00150>.
- (147) Madsen, A. S.; Andersen, C.; Daoud, M.; Anderson, K. A.; Laursen, J. S.; Chakladar, S.; Huynh, F. K.; Colaço, A. R.; Backos, D. S.; Fristrup, P.; Hirschey, M. D.; Olsen, C. A. Investigating the Sensitivity of NAD⁺-Dependent Sirtuin Deacylation Activities to NADH *. *Journal of Biological Chemistry* **2016**, *291* (13), 7128–7141. <https://doi.org/10.1074/jbc.M115.668699>.
- (148) Teng, Y.-B.; Jing, H.; Aramsangtienchai, P.; He, B.; Khan, S.; Hu, J.; Lin, H.; Hao, Q. Efficient Demyristoylase Activity of SIRT2 Revealed by Kinetic and Structural Studies. *Scientific Reports* **2015**, *5* (1), 8529. <https://doi.org/10.1038/srep08529>.
- (149) Huang, H.; Zhang, D.; Wang, Y.; Perez-Neut, M.; Han, Z.; Zheng, Y. G.; Hao, Q.; Zhao, Y. Lysine Benzoylation Is a Histone Mark Regulated by SIRT2. *Nat Commun* **2018**, *9* (1), 3374. <https://doi.org/10.1038/s41467-018-05567-w>.
- (150) Bi, D.; Yang, J.; Hong, J. Y.; Parikh, P.; Hinds, N.; Infanti, J.; Lin, H.; Weiser, B. P. Substrate-Dependent Modulation of SIRT2 by a Fluorescent Probe, 1-Aminoanthracene. *Biochemistry* **2020**, *59* (40), 3869–3878. <https://doi.org/10.1021/acs.biochem.0c00564>.
- (151) Yang, W.; Chen, W.; Su, H.; Li, R.; Song, C.; Wang, Z.; Yang, L. Recent Advances in the Development of Histone Deacylase SIRT2 Inhibitors. *RSC Advances* **2020**, *10* (61), 37382–37390. <https://doi.org/10.1039/D0RA06316A>.
- (152) Vaquero, A.; Scher, M. B.; Lee, D. H.; Sutton, A.; Cheng, H.-L.; Alt, F. W.; Serrano, L.; Sternglanz, R.; Reinberg, D. SirT2 Is a Histone Deacetylase with Preference for Histone H4 Lys 16 during Mitosis. *Genes Dev* **2006**, *20* (10), 1256–1261. <https://doi.org/10.1101/gad.1412706>.
- (153) Wang, F.; Nguyen, M.; Qin, F. X.-F.; Tong, Q. SIRT2 Deacetylates FOXO3a in Response to Oxidative Stress and Caloric Restriction. *Aging Cell* **2007**, *6* (4), 505–514. <https://doi.org/10.1111/j.1474-9726.2007.00304.x>.
- (154) Jing, E.; Gesta, S.; Kahn, C. R. SIRT2 Regulates Adipocyte Differentiation through FoxO1 Acetylation/Deacetylation. *Cell Metab* **2007**, *6* (2), 105–114. <https://doi.org/10.1016/j.cmet.2007.07.003>.
- (155) Zhao, D.; Zou, S.-W.; Liu, Y.; Zhou, X.; Mo, Y.; Wang, P.; Xu, Y.-H.; Dong, B.; Xiong, Y.; Lei, Q.-Y.; Guan, K.-L. Lysine-5 Acetylation Negatively Regulates Lactate Dehydrogenase

- A and Is Decreased in Pancreatic Cancer. *Cancer Cell* **2013**, *23* (4), 464–476. <https://doi.org/10.1016/j.ccr.2013.02.005>.
- (156) Park, S.-H.; Ozden, O.; Liu, G.; Song, H. Y.; Zhu, Y.; Yan, Y.; Zou, X.; Kang, H.-J.; Jiang, H.; Principe, D. R.; Cha, Y.-I.; Roh, M.; Vassilopoulos, A.; Gius, D. SIRT2-Mediated Deacetylation and Tetramerization of Pyruvate Kinase Directs Glycolysis and Tumor Growth. *Cancer Res* **2016**, *76* (13), 3802–3812. <https://doi.org/10.1158/0008-5472.CAN-15-2498>.
- (157) Machado de Oliveira, Ri.; Sarkander, J.; Kazantsev, A.; Outeiro, T. SIRT2 as a Therapeutic Target for Age-Related Disorders. *Frontiers in Pharmacology* **2012**, *3*.
- (158) Zhang, L.; Kim, S.; Ren, X. The Clinical Significance of SIRT2 in Malignancies: A Tumor Suppressor or an Oncogene? *Front. Oncol.* **2020**, *10*. <https://doi.org/10.3389/fonc.2020.01721>.
- (159) Hong, J. Y.; Lin, H. Sirtuin Modulators in Cellular and Animal Models of Human Diseases. *Front Pharmacol* **2021**, *12*, 735044. <https://doi.org/10.3389/fphar.2021.735044>.
- (160) Carafa, V.; Nebbioso, A.; Cuomo, F.; Rotili, D.; Cobellis, G.; Bontempo, P.; Baldi, A.; Spugnini, E. P.; Citro, G.; Chambery, A.; Russo, R.; Ruvo, M.; Ciana, P.; Maravigna, L.; Shaik, J.; Radaelli, E.; Antonellis, P. D.; Tarantino, D.; Pirolli, A.; Ragno, R.; Zollo, M.; Stunnenberg, H. G.; Mai, A.; Altucci, L. RIP1–HAT1–SIRT Complex Identification and Targeting in Treatment and Prevention of Cancer. *Clin Cancer Res* **2018**, *24* (12), 2886–2900. <https://doi.org/10.1158/1078-0432.CCR-17-3081>.
- (161) Yi, F.; Zhang, Y.; Wang, Z.; Wang, Z.; Li, Z.; Zhou, T.; Xu, H.; Liu, J.; Jiang, B.; Li, X.; Wang, L.; Bai, N.; Guo, Q.; Guan, Y.; Feng, Y.; Mao, Z.; Fan, G.; Zhang, S.; Wang, C.; Cao, L.; O'Rourke, B. P.; Wang, Y.; Wu, Y.; Wu, B.; You, S.; Zhang, N.; Guan, J.; Song, X.; Sun, Y.; Wei, S.; Cao, L. The Deacetylation-Phosphorylation Regulation of SIRT2-SMC1A Axis as a Mechanism of Antimitotic Catastrophe in Early Tumorigenesis. *Science Advances* **2021**, *7* (9), eabe5518. <https://doi.org/10.1126/sciadv.abe5518>.
- (162) Jing, H.; Hu, J.; He, B.; Negron Abril, Y. L.; Stupinski, J.; Weiser, K.; Carbonaro, M.; Chiang, Y.-L.; Southard, T.; Giannakakou, P.; Weiss, R. S.; Lin, H. A SIRT2-Selective Inhibitor Promotes c-Myc Oncoprotein Degradation and Exhibits Broad Anticancer Activity. *Cancer Cell* **2016**, *29* (3), 297–310. <https://doi.org/10.1016/j.ccell.2016.02.007>.
- (163) Nielsen, A. L.; Rajabi, N.; Kudo, N.; Lundø, K.; Moreno-Yruela, C.; Bæk, M.; Fontenas, M.; Lucidi, A.; Madsen, A. S.; Yoshida, M.; Olsen, C. A. Mechanism-Based Inhibitors of SIRT2: Structure–Activity Relationship, X-Ray Structures, Target Engagement, Regulation of α -Tubulin Acetylation and Inhibition of Breast Cancer Cell Migration. *RSC Chem. Biol.* **2021**. <https://doi.org/10.1039/D0CB00036A>.
- (164) Wawruszak, A.; Luszczki, J.; Czerwonka, A.; Okon, E.; Stepulak, A. Assessment of Pharmacological Interactions between SIRT2 Inhibitor AGK2 and Paclitaxel in Different Molecular Subtypes of Breast Cancer Cells. *Cells* **2022**, *11* (7), 1211. <https://doi.org/10.3390/cells11071211>.
- (165) Karwaciak, I.; Sałkowska, A.; Karaś, K.; Sobalska-Kwapis, M.; Walczak-Drzewiecka, A.; Pułaski, Ł.; Strapagiel, D.; Dastyk, J.; Ratajewski, M. SIRT2 Contributes to the Resistance of Melanoma Cells to the Multikinase Inhibitor Dasatinib. *Cancers (Basel)* **2019**, *11* (5), 673. <https://doi.org/10.3390/cancers11050673>.
- (166) Cheng, W.-L.; Chen, K.-Y.; Lee, K.-Y.; Feng, P.-H.; Wu, S.-M. Nicotinic-NAChR Signaling Mediates Drug Resistance in Lung Cancer. *J. Cancer* **2020**, *11* (5), 1125–1140. <https://doi.org/10.7150/jca.36359>.
- (167) Hamaidi, I.; Zhang, L.; Kim, N.; Wang, M.-H.; Iclozan, C.; Fang, B.; Liu, M.; Koomen, J. M.; Berglund, A. E.; Yoder, S. J.; Yao, J.; Engelman, R. W.; Creelan, B. C.; Conejo-Garcia, J. R.; Antonia, S. J.; Mulé, J. J.; Kim, S. Sirt2 Inhibition Enhances Metabolic Fitness and Effector Functions of Tumor-Reactive T Cells. *Cell Metabolism* **2020**, *32* (3), 420–436.e12. <https://doi.org/10.1016/j.cmet.2020.07.008>.

- (168) Carafa, V.; Altucci, L.; Nebbioso, A. Dual Tumor Suppressor and Tumor Promoter Action of Sirtuins in Determining Malignant Phenotype. *Front Pharmacol* **2019**, *10*, 38. <https://doi.org/10.3389/fphar.2019.00038>.
- (169) Yin, L.; Duan, J.-J.; Bian, X.-W.; Yu, S. Triple-Negative Breast Cancer Molecular Subtyping and Treatment Progress. *Breast Cancer Research* **2020**, *22* (1), 61. <https://doi.org/10.1186/s13058-020-01296-5>.
- (170) Zhou, W.; Ni, T. K.; Wronski, A.; Glass, B.; Skibinski, A.; Beck, A.; Kuperwasser, C. The SIRT2 Deacetylase Stabilizes Slug to Control Malignancy of Basal-like Breast Cancer. *Cell Rep* **2016**, *17* (5), 1302–1317. <https://doi.org/10.1016/j.celrep.2016.10.006>.
- (171) Ahmadi, S. E.; Rahimi, S.; Zarandi, B.; Chegeni, R.; Safa, M. MYC: A Multipurpose Oncogene with Prognostic and Therapeutic Implications in Blood Malignancies. *Journal of Hematology & Oncology* **2021**, *14* (1), 121. <https://doi.org/10.1186/s13045-021-01111-4>.
- (172) Amjadi-Moheb, F.; Paniri, A.; Akhavan-Niaki, H. Insights into the Links between MYC and 3D Chromatin Structure and Epigenetics Regulation: Implications for Cancer Therapy. *Cancer Res* **2021**, *81* (8), 1925–1936. <https://doi.org/10.1158/0008-5472.CAN-20-3613>.
- (173) Poole, C. J.; van Riggelen, J. MYC—Master Regulator of the Cancer Epigenome and Transcriptome. *Genes (Basel)* **2017**, *8* (5), 142. <https://doi.org/10.3390/genes8050142>.
- (174) Thomas, L. R.; Adams, C. M.; Fesik, S. W.; Eischen, C. M.; Tansey, W. P. Targeting MYC through WDR5. *Molecular & Cellular Oncology* **2020**, *7* (2), 1709388. <https://doi.org/10.1080/23723556.2019.1709388>.
- (175) Thomas, L. R.; Adams, C. M.; Wang, J.; Weissmiller, A. M.; Creighton, J.; Lorey, S. L.; Liu, Q.; Fesik, S. W.; Eischen, C. M.; Tansey, W. P. Interaction of the Oncoprotein Transcription Factor MYC with Its Chromatin Cofactor WDR5 Is Essential for Tumor Maintenance. *PNAS* **2019**, *116* (50), 25260–25268. <https://doi.org/10.1073/pnas.1910391116>.
- (176) Ružić, D.; Đoković, N.; Nikolić, K.; Vujić, Z. Medicinal Chemistry of Histone Deacetylase Inhibitors. *Arhiv za farmaciju* **2021**, *71* (2), 73–100. <https://doi.org/10.5937/arhfarm71-30618>.
- (177) Sauve, A. A.; Youn, D. Y. Sirtuins: NAD⁺-Dependent Deacetylase Mechanism and Regulation. *Current Opinion in Chemical Biology* **2012**, *16* (5), 535–543. <https://doi.org/10.1016/j.cbpa.2012.10.003>.
- (178) Finnin, M. S.; Donigian, J. R.; Pavletich, N. P. Structure of the Histone Deacetylase SIRT2. *Nat Struct Biol* **2001**, *8* (7), 621–625. <https://doi.org/10.1038/89668>.
- (179) Robaa, D.; Monaldi, D.; Wössner, N.; Kudo, N.; Rumpf, T.; Schiedel, M.; Yoshida, M.; Jung, M. Opening the Selectivity Pocket in the Human Lysine Deacetylase Sirtuin2 - New Opportunities, New Questions. *Chem Rec* **2018**, *18* (12), 1701–1707. <https://doi.org/10.1002/tcr.201800044>.
- (180) Rumpf, T.; Schiedel, M.; Karaman, B.; Roessler, C.; North, B. J.; Lehotzky, A.; Oláh, J.; Ladwein, K. I.; Schmidtkunz, K.; Gajer, M.; Pannek, M.; Steegborn, C.; Sinclair, D. A.; Gerhardt, S.; Ovádi, J.; Schutkowski, M.; Sippl, W.; Einsle, O.; Jung, M. Selective Sirt2 Inhibition by Ligand-Induced Rearrangement of the Active Site. *Nature Communications* **2015**, *6* (1), 6263. <https://doi.org/10.1038/ncomms7263>.
- (181) Wang, Y.; Fung, Y. M. E.; Zhang, W.; He, B.; Chung, M. W. H.; Jin, J.; Hu, J.; Lin, H.; Hao, Q. De-Acylation Mechanism by SIRT2 Revealed in the 1'-SH-2'-O-Myristoyl Intermediate Structure. *Cell Chem Biol* **2017**, *24* (3), 339–345. <https://doi.org/10.1016/j.chembiol.2017.02.007>.
- (182) Moniot, S.; Schutkowski, M.; Steegborn, C. Crystal Structure Analysis of Human Sirt2 and Its ADP-Ribose Complex. *Journal of Structural Biology* **2013**, *182* (2), 136–143. <https://doi.org/10.1016/j.jsb.2013.02.012>.
- (183) Wang, Y.; He, J.; Liao, M.; Hu, M.; Li, W.; Ouyang, H.; Wang, X.; Ye, T.; Zhang, Y.; Ouyang, L. An Overview of Sirtuins as Potential Therapeutic Target: Structure, Function and Modulators. *European Journal of Medicinal Chemistry* **2019**, *161*, 48–77. <https://doi.org/10.1016/j.ejmech.2018.10.028>.

- (184) Hong, J. Y.; Fernandez, I.; Anmangandla, A.; Lu, X.; Bai, J. J.; Lin, H. Pharmacological Advantage of SIRT2-Selective versus Pan-SIRT1-3 Inhibitors. *ACS Chem Biol* **2021**, *16* (7), 1266–1275. <https://doi.org/10.1021/acscchembio.1c00331>.
- (185) Sakkiah, S.; Arooj, M.; Cao, G. P.; Lee, K. W. Insight the C-Site Pocket Conformational Changes Responsible for Sirtuin 2 Activity Using Molecular Dynamics Simulations. *PLoS ONE* **2013**, *8* (3), e59278. <https://doi.org/10.1371/journal.pone.0059278>.
- (186) Singh, M.; Srivastava, M.; Wakode, S. R.; Asthana, S. Elucidation of Structural Determinants Delineates the Residues Playing Key Roles in Differential Dynamics and Selective Inhibition of Sirt1–3. *J. Chem. Inf. Model.* **2021**, *61* (3), 1105–1124. <https://doi.org/10.1021/acs.jcim.0c01193>.
- (187) Muvva, C.; Murugan, N. A.; Kumar Choutipalli, V. S.; Subramanian, V. Unraveling the Unbinding Pathways of Products Formed in Catalytic Reactions Involved in SIRT1–3: A Random Acceleration Molecular Dynamics Simulation Study. *J. Chem. Inf. Model.* **2019**, *59* (10), 4100–4115. <https://doi.org/10.1021/acs.jcim.9b00513>.
- (188) Kiviranta, P. H.; Salo, H. S.; Leppänen, J.; Rinne, V. M.; Kyrlylenko, S.; Kuusisto, E.; Suuronen, T.; Salminen, A.; Poso, A.; Lahtela-Kakkonen, M.; Wallén, E. A. A. Characterization of the Binding Properties of SIRT2 Inhibitors with a N-(3-Phenylpropenoyl)-Glycine Tryptamide Backbone. *Bioorganic & Medicinal Chemistry* **2008**, *16* (17), 8054–8062. <https://doi.org/10.1016/j.bmc.2008.07.059>.
- (189) Bryan, A. F.; Wang, J.; Howard, G. C.; Guarnaccia, A. D.; Woodley, C. M.; Aho, E. R.; Rellinger, E. J.; Matlock, B. K.; Flaherty, D. K.; Lorey, S. L.; Chung, D. H.; Fesik, S. W.; Liu, Q.; Weissmiller, A. M.; Tansey, W. P. WDR5 Is a Conserved Regulator of Protein Synthesis Gene Expression. *Nucleic Acids Research* **2020**, *48* (6), 2924–2941. <https://doi.org/10.1093/nar/gkaa051>.
- (190) Guarnaccia, A. D.; Rose, K. L.; Wang, J.; Zhao, B.; Popay, T. M.; Wang, C. E.; Guerrazzi, K.; Hill, S.; Woodley, C. M.; Hansen, T. J.; Lorey, S. L.; Shaw, J. G.; Payne, W. G.; Weissmiller, A. M.; Olejniczak, E. T.; Fesik, S. W.; Liu, Q.; Tansey, W. P. Impact of WIN Site Inhibitor on the WDR5 Interactome. *Cell Rep* **2021**, *34* (3), 108636. <https://doi.org/10.1016/j.celrep.2020.108636>.
- (191) Thomas, L. R.; Wang, Q.; Grieb, B. C.; Phan, J.; Foshage, A. M.; Sun, Q.; Olejniczak, E. T.; Clark, T.; Dey, S.; Lorey, S.; Alicie, B.; Howard, G. C.; Cawthon, B.; Ess, K. C.; Eischen, C. M.; Zhao, Z.; Fesik, S. W.; Tansey, W. P. Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Mol Cell* **2015**, *58* (3), 440–452. <https://doi.org/10.1016/j.molcel.2015.02.028>.
- (192) Yu, X.; Li, D.; Kottur, J.; Shen, Y.; Kim, H. S.; Park, K.-S.; Tsai, Y.-H.; Gong, W.; Wang, J.; Suzuki, K.; Parker, J.; Herring, L.; Kaniskan, H. Ü.; Cai, L.; Jain, R.; Liu, J.; Aggarwal, A. K.; Wang, G. G.; Jin, J. A Selective WDR5 Degradator Inhibits Acute Myeloid Leukemia in Patient-Derived Mouse Models. *Sci Transl Med* **2021**, *13* (613), eabj1578. <https://doi.org/10.1126/scitranslmed.abj1578>.
- (193) Chen, W.-L.; Li, D.-D.; Chen, X.; Wang, Y.-Z.; Xu, J.-J.; Jiang, Z.-Y.; You, Q.-D.; Guo, X.-K. Proton Pump Inhibitors Selectively Suppress MLL Rearranged Leukemia Cells via Disrupting MLL1-WDR5 Protein-Protein Interaction. *Eur J Med Chem* **2020**, *188*, 112027. <https://doi.org/10.1016/j.ejmech.2019.112027>.
- (194) Milne, T. A.; Martin, M. E.; Brock, H. W.; Slany, R. K.; Hess, J. L. Leukemogenic MLL Fusion Proteins Bind across a Broad Region of the Hox A9 Locus, Promoting Transcription and Multiple Histone Modifications. *Cancer Res* **2005**, *65* (24), 11367–11374. <https://doi.org/10.1158/0008-5472.CAN-05-1041>.
- (195) Teuscher, K. B.; Chowdhury, S.; Meyers, K. M.; Tian, J.; Sai, J.; Van Meveren, M.; South, T. M.; Sensintaffar, J. L.; Rietz, T. A.; Goswami, S.; Wang, J.; Grieb, B. C.; Lorey, S. L.; Howard, G. C.; Liu, Q.; Moore, W. J.; Stott, G. M.; Tansey, W. P.; Lee, T.; Fesik, S. W. Structure-Based Discovery of Potent WD Repeat Domain 5 Inhibitors That Demonstrate

- Efficacy and Safety in Preclinical Animal Models. *Proceedings of the National Academy of Sciences* **2023**, *120* (1), e2211297120. <https://doi.org/10.1073/pnas.2211297120>.
- (196) WDR5 Is a MYC Cofactor Required for MYC-Driven Tumorigenesis. *Cancer Discovery* **2015**, *5* (5), 461. <https://doi.org/10.1158/2159-8290.CD-RW2015-062>.
- (197) Odho, Z.; Southall, S. M.; Wilson, J. R. Characterization of a Novel WDR5-Binding Site That Recruits RbBP5 through a Conserved Motif to Enhance Methylation of Histone H3 Lysine 4 by Mixed Lineage Leukemia Protein-1. *J Biol Chem* **2010**, *285* (43), 32967–32976. <https://doi.org/10.1074/jbc.M110.159921>.
- (198) Zhang, P.; Lee, H.; Brunzelle, J. S.; Couture, J.-F. The Plasticity of WDR5 Peptide-Binding Cleft Enables the Binding of the SET1 Family of Histone Methyltransferases. *Nucleic Acids Research* **2012**, *40* (9), 4237–4246. <https://doi.org/10.1093/nar/gkr1235>.
- (199) Park, S. H.; Ayoub, A.; Lee, Y.-T.; Xu, J.; Kim, H.; Zheng, W.; Zhang, B.; Sha, L.; An, S.; Zhang, Y.; Cianfrocco, M. A.; Su, M.; Dou, Y.; Cho, U.-S. Cryo-EM Structure of the Human MLL1 Core Complex Bound to the Nucleosome. *Nat Commun* **2019**, *10* (1), 5540. <https://doi.org/10.1038/s41467-019-13550-2>.
- (200) Cai, W. L.; Chen, J. F.-Y.; Chen, H.; Wingrove, E.; Kurley, S. J.; Chan, L. H.; Zhang, M.; Arnal-Estape, A.; Zhao, M.; Balabaki, A.; Li, W.; Yu, X.; Krop, E. D.; Dou, Y.; Liu, Y.; Jin, J.; Westbrook, T. F.; Nguyen, D. X.; Yan, Q. Human WDR5 Promotes Breast Cancer Growth and Metastasis via KMT2-Independent Translation Regulation. *eLife* **2022**, *11*, e78163. <https://doi.org/10.7554/eLife.78163>.
- (201) Macalino, S. J. Y.; Basith, S.; Clavio, N. A. B.; Chang, H.; Kang, S.; Choi, S. Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery. *Molecules* **2018**, *23* (8). <https://doi.org/10.3390/molecules23081963>.
- (202) Arkin, M. R.; Tang, Y.; Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chemistry & Biology* **2014**, *21* (9), 1102–1114. <https://doi.org/10.1016/j.chembiol.2014.09.001>.
- (203) Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small Molecules, Big Targets: Drug Discovery Faces the Protein–Protein Interaction Challenge. *Nature Reviews Drug Discovery* **2016**, *15* (8), 533–550. <https://doi.org/10.1038/nrd.2016.29>.
- (204) Ran, X.; Gestwicki, J. E. Inhibitors of Protein-Protein Interactions (PPIs): An Analysis of Scaffold Choices and Buried Surface Area. *Curr Opin Chem Biol* **2018**, *44*, 75–86. <https://doi.org/10.1016/j.cbpa.2018.06.004>.
- (205) Shin, W.-H.; Christoffer, C. W.; Kihara, D. In Silico Structure-Based Approaches to Discover Protein-Protein Interaction-Targeting Drugs. *Methods* **2017**, *131*, 22–32. <https://doi.org/10.1016/j.jymeth.2017.08.006>.
- (206) Dai, X.; Guo, W.; Zhan, C.; Liu, X.; Bai, Z.; Yang, Y. WDR5 Expression Is Prognostic of Breast Cancer Outcome. *PLoS One* **2015**, *10* (9), e0124964. <https://doi.org/10.1371/journal.pone.0124964>.
- (207) Ge, Z.; Song, E. J.; Kawasawa, Y. I.; Li, J.; Dovat, S.; Song, C. WDR5 High Expression and Its Effect on Tumorigenesis in Leukemia. *Oncotarget* **2016**, *7* (25), 37740–37754. <https://doi.org/10.18632/oncotarget.9312>.
- (208) Cui, Z.; Li, H.; Liang, F.; Mu, C.; Mu, Y.; Zhang, X.; Liu, J. Effect of High WDR5 Expression on the Hepatocellular Carcinoma Prognosis. *Oncol Lett* **2018**, *15* (5), 7864–7870. <https://doi.org/10.3892/ol.2018.8298>.
- (209) Sun, W.; Guo, F.; Liu, M. Up-Regulated WDR5 Promotes Gastric Cancer Formation by Induced Cyclin D1 Expression. *J Cell Biochem* **2018**, *119* (4), 3304–3316. <https://doi.org/10.1002/jcb.26491>.
- (210) Wu, Y.; Diao, P.; Li, Z.; Zhang, W.; Wang, D.; Wang, Y.; Cheng, J. Overexpression of WD Repeat Domain 5 Associates with Aggressive Clinicopathological Features and Unfavorable Prognosis in Head Neck Squamous Cell Carcinoma. *J Oral Pathol Med* **2018**, *47* (5), 502–510. <https://doi.org/10.1111/jop.12708>.

- (211) Punzi, S.; Balestrieri, C.; D'Alesio, C.; Bossi, D.; Dellino, G. I.; Gatti, E.; Pruneri, G.; Criscitiello, C.; Lovati, G.; Meliksetyan, M.; Carugo, A.; Curigliano, G.; Natoli, G.; Pelicci, P. G.; Lanfrancone, L. WDR5 Inhibition Halts Metastasis Dissemination by Repressing the Mesenchymal Phenotype of Breast Cancer Cells. *Breast Cancer Research* **2019**, *21* (1), 123. <https://doi.org/10.1186/s13058-019-1216-y>.
- (212) Albert, L.; Xu, J.; Wan, R.; Srinivasan, V.; Dou, Y.; Vázquez, O. Controlled Inhibition of Methyltransferases Using Photoswitchable Peptidomimetics: Towards an Epigenetic Regulation of Leukemia. *Chem. Sci.* **2017**, *8* (6), 4612–4618. <https://doi.org/10.1039/C7SC00137A>.
- (213) Albert, L.; Peñalver, A.; Djokovic, N.; Werel, L.; Hoffarth, M.; Ruzic, D.; Xu, J.; Essen, L.-O.; Nikolic, K.; Dou, Y.; Vázquez, O. Modulating Protein–Protein Interactions with Visible-Light-Responsive Peptide Backbone Photoswitches. *ChemBioChem* **2019**, *20* (11), 1417–1429. <https://doi.org/10.1002/cbic.201800737>.
- (214) Albert, L.; Nagpal, J.; Steinchen, W.; Zhang, L.; Werel, L.; Djokovic, N.; Ruzic, D.; Hoffarth, M.; Xu, J.; Kaspereit, J.; Abendroth, F.; Royant, A.; Bange, G.; Nikolic, K.; Ryu, S.; Dou, Y.; Essen, L.-O.; Vázquez, O. Bistable Photoswitch Allows in Vivo Control of Hematopoiesis. *ACS Cent. Sci.* **2022**, *8* (1), 57–66. <https://doi.org/10.1021/acscentsci.1c00434>.
- (215) Kobauri, P.; Dekker, F. J.; Szymanski, W.; Feringa, B. L. Rational Design in Photopharmacology with Molecular Photoswitches. *Angewandte Chemie International Edition n/a* (n/a), e202300681. <https://doi.org/10.1002/anie.202300681>.
- (216) Gauthier, J.; Vincent, A. T.; Charette, S. J.; Derome, N. A Brief History of Bioinformatics. *Briefings in Bioinformatics* **2019**, *20* (6), 1981–1996. <https://doi.org/10.1093/bib/bby063>.
- (217) Fernald, G. H.; Capriotti, E.; Daneshjou, R.; Karczewski, K. J.; Altman, R. B. Bioinformatics Challenges for Personalized Medicine. *Bioinformatics* **2011**, *27* (13), 1741–1748. <https://doi.org/10.1093/bioinformatics/btr295>.
- (218) Azuaje, F. *Bioinformatics and Biomarker Discovery: “Omic” Data Analysis for Personalized Medicine*; John Wiley & Sons, 2011.
- (219) You, Y.; Lai, X.; Pan, Y.; Zheng, H.; Vera, J.; Liu, S.; Deng, S.; Zhang, L. Artificial Intelligence in Cancer Target Identification and Drug Discovery. *Sig Transduct Target Ther* **2022**, *7* (1), 1–24. <https://doi.org/10.1038/s41392-022-00994-0>.
- (220) Ramsay, R. R.; Popovic-Nikolic, M. R.; Nikolic, K.; Uliassi, E.; Bolognesi, M. L. A Perspective on Multi-Target Drug Discovery and Design for Complex Diseases. *Clin Transl Med* **2018**, *7*, 3. <https://doi.org/10.1186/s40169-017-0181-2>.
- (221) Creighton, C. J. Making Use of Cancer Genomic Databases. *Curr Protoc Mol Biol* **2018**, *121*, 19.14.1–19.14.13. <https://doi.org/10.1002/cpmb.49>.
- (222) Rees, M. G.; Seashore-Ludlow, B.; Cheah, J. H.; Adams, D. J.; Price, E. V.; Gill, S.; Javaid, S.; Coletti, M. E.; Jones, V. L.; Bodycombe, N. E.; Soule, C. K.; Alexander, B.; Li, A.; Montgomery, P.; Kotz, J. D.; Hon, C. S.-Y.; Munoz, B.; Liefeld, T.; Dančik, V.; Haber, D. A.; Clish, C. B.; Bittker, J. A.; Palmer, M.; Wagner, B. K.; Clemons, P. A.; Shamji, A. F.; Schreiber, S. L. Correlating Chemical Sensitivity and Basal Gene Expression Reveals Mechanism of Action. *Nat Chem Biol* **2016**, *12* (2), 109–116. <https://doi.org/10.1038/nchembio.1986>.
- (223) Yang, H.; Sun, B.; Xu, K.; He, Y.; Zhang, T.; Hall, S. R. R.; Tan, S. T.; Schmid, R. A.; Peng, R.-W.; Hu, G.; Yao, F. Pharmaco-Transcriptomic Correlation Analysis Reveals Novel Responsive Signatures to HDAC Inhibitors and Identifies Dasatinib as a Synergistic Interactor in Small-Cell Lung Cancer. *eBioMedicine* **2021**, *69*, 103457. <https://doi.org/10.1016/j.ebiom.2021.103457>.
- (224) She, S.; Chen, H.; Ji, W.; Sun, M.; Cheng, J.; Rui, M.; Feng, C. Deep Learning-Based Multi-Drug Synergy Prediction Model for Individually Tailored Anti-Cancer Therapies. *Frontiers in Pharmacology* **2022**, *13*.

- (225) Li, M.; Wang, Y.; Zheng, R.; Shi, X.; Li, Y.; Wu, F.-X.; Wang, J. DeepDSC: A Deep Learning Method to Predict Drug Sensitivity of Cancer Cell Lines. *IEEE/ACM Trans Comput Biol Bioinform* **2021**, *18* (2), 575–582. <https://doi.org/10.1109/TCBB.2019.2919581>.
- (226) Sundriyal, S.; Moniot, S.; Mahmud, Z.; Yao, S.; Di Fruscia, P.; Reynolds, C. R.; Dexter, D. T.; Sternberg, M. J. E.; Lam, E. W.-F.; Steegborn, C.; Fuchter, M. J. Thienopyrimidinone Based Sirtuin-2 (SIRT2)-Selective Inhibitors Bind in the Ligand Induced Selectivity Pocket. *J Med Chem* **2017**, *60* (5), 1928–1945. <https://doi.org/10.1021/acs.jmedchem.6b01690>.
- (227) Mellini, P.; Itoh, Y.; Tsumoto, H.; Li, Y.; Suzuki, M.; Tokuda, N.; Kakizawa, T.; Miura, Y.; Takeuchi, J.; Lahtela-Kakkonen, M.; Suzuki, T. Potent Mechanism-Based Sirtuin-2-Selective Inhibition by an in Situ-Generated Occupant of the Substrate-Binding Site, “Selectivity Pocket” and NAD⁺-Binding Site. *Chem. Sci.* **2017**, *8* (9), 6400–6408. <https://doi.org/10.1039/C7SC02738A>.
- (228) Yang, L.-L.; Wang, H.-L.; Zhong, L.; Yuan, C.; Liu, S.-Y.; Yu, Z.-J.; Liu, S.; Yan, Y.-H.; Wu, C.; Wang, Y.; Wang, Z.; Yu, Y.; Chen, Q.; Li, G.-B. X-Ray Crystal Structure Guided Discovery of New Selective, Substrate-Mimicking Sirtuin 2 Inhibitors That Exhibit Activities against Non-Small Cell Lung Cancer Cells. *European Journal of Medicinal Chemistry* **2018**, *155*, 806–823. <https://doi.org/10.1016/j.ejmech.2018.06.041>.
- (229) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52* (12), 3155–3168. <https://doi.org/10.1021/ci3003649>.
- (230) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52* (12), 3144–3154. <https://doi.org/10.1021/ci300363c>.
- (231) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 33–38, 27–28. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (232) Neese, F. Software Update: The ORCA Program System, Version 4.0. *WIREs Computational Molecular Science* **2018**, *8* (1), e1327. <https://doi.org/10.1002/wcms.1327>.
- (233) Neese, F. The ORCA Program System. *WIREs Computational Molecular Science* **2012**, *2* (1), 73–78. <https://doi.org/10.1002/wcms.81>.
- (234) Giannos, T.; Lešnik, S.; Bren, U.; Hodošček, M.; Domratcheva, T.; Bondar, A.-N. CHARMM Force-Field Parameters for Morphine, Heroin, and Oliceridine, and Conformational Dynamics of Opioid Drugs. *J. Chem. Inf. Model.* **2021**, *61* (8), 3964–3977. <https://doi.org/10.1021/acs.jcim.1c00667>.
- (235) Vermaas, J. V.; Petridis, L.; Ralph, J.; Crowley, M. F.; Beckham, G. T. Systematic Parameterization of Lignin for the CHARMM Force Field. *Green Chem.* **2019**, *21* (1), 109–122. <https://doi.org/10.1039/C8GC03209B>.
- (236) Baday, S. Optimization of CHARMM Force Field Parameters of a Prostate Cancer Drug Enzalutamide. *Journal of Molecular Graphics and Modelling* **2020**, *96*, 107541. <https://doi.org/10.1016/j.jmgm.2020.107541>.
- (237) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. CHARMM General Force Field (CGenFF): A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem* **2010**, *31* (4), 671–690. <https://doi.org/10.1002/jcc.21367>.
- (238) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* **1993**, *234* (3), 779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
- (239) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W522-525. <https://doi.org/10.1093/nar/gkm276>.

- (240) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J Chem Theory Comput* **2011**, *7* (2), 525–537. <https://doi.org/10.1021/ct100578z>.
- (241) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PKa Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295. <https://doi.org/10.1021/ct200133y>.
- (242) Marvin 21.3.0, 2021.
- (243) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (244) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J Comput Chem* **2013**, *34* (25), 2135–2145. <https://doi.org/10.1002/jcc.23354>.
- (245) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- (246) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. <https://doi.org/10.1002/jcc.21787>.
- (247) Serçinoglu, O.; Ozbek, P. GRINN: A Tool for Calculation of Residue Interaction Energies and Protein Energy Network Analysis of Molecular Dynamics Simulations. *Nucleic Acids Res.* **2018**, *46* (W1), W554–W562. <https://doi.org/10.1093/nar/gky381>.
- (248) Goenawan, I. H.; Bryan, K.; Lynn, D. J. DyNet: Visualization and Analysis of Dynamic Molecular Interaction Networks. *Bioinformatics* **2016**, *32* (17), 2713–2715. <https://doi.org/10.1093/bioinformatics/btw187>.
- (249) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- (250) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Gunsteren, W. F. van; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition* **1999**, *38* (1–2), 236–240. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990115\)38:1/2<236::AID-ANIE236>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M).
- (251) Molgedey, L.; Schuster, H. G. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.* **1994**, *72* (23), 3634–3637. <https://doi.org/10.1103/PhysRevLett.72.3634>.
- (252) Methods Using Time Structure. In *Independent Component Analysis*; John Wiley & Sons, Ltd, 2001; pp 341–354. <https://doi.org/10.1002/0471221317.ch18>.
- (253) Schultze, S.; Grubmüller, H. Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. *J. Chem. Theory Comput.* **2021**, *17* (9), 5766–5776. <https://doi.org/10.1021/acs.jctc.1c00273>.
- (254) M. Sultan, M.; Pande, V. S. TICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017**, *13* (6), 2440–2447. <https://doi.org/10.1021/acs.jctc.7b00182>.
- (255) Harrigan, M. P.; Sultan, M. M.; Hernández, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; McGibbon, R. T.; Pande, V. S. MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophysical Journal* **2017**, *112* (1), 10–15. <https://doi.org/10.1016/j.bpj.2016.10.042>.
- (256) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.;

- Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A.; The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nature Methods* **2019**, *16* (8), 670–673. <https://doi.org/10.1038/s41592-019-0506-8>.
- (257) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Computer Physics Communications* **2014**, *185* (2), 604–613. <https://doi.org/10.1016/j.cpc.2013.09.018>.
- (258) Zhang, H.; Kai, E. L. J.; Lu, L. Investigating the Stability of Dengue Virus Envelope Protein Dimer Using Well-Tempered Metadynamics Simulations. *Proteins: Structure, Function, and Bioinformatics* **2020**, *88* (5), 643–653. <https://doi.org/10.1002/prot.25844>.
- (259) Incerti, M.; Russo, S.; Callegari, D.; Pala, D.; Giorgio, C.; Zanotti, I.; Barocelli, E.; Vicini, P.; Vacondio, F.; Rivara, S.; Castelli, R.; Tognolini, M.; Lodola, A. Metadynamics for Perspective Drug Design: Computationally Driven Synthesis of New Protein–Protein Interaction Inhibitors Targeting the EphA2 Receptor. *J. Med. Chem.* **2017**, *60* (2), 787–796. <https://doi.org/10.1021/acs.jmedchem.6b01642>.
- (260) Fidelak, J.; Juraszek, J.; Branduardi, D.; Bianciotto, M.; Gervasio, F. L. *Free-Energy-Based Methods for Binding Profile Determination in a Congeneric Series of CDK2 Inhibitors*. ACS Publications. <https://doi.org/10.1021/jp911689r>.
- (261) Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T. An Efficient Metadynamics-Based Protocol To Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *J. Chem. Inf. Model.* **2017**, *57* (5), 1210–1217. <https://doi.org/10.1021/acs.jcim.6b00772>.
- (262) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47* (2), 279–294. <https://doi.org/10.1021/ci600253e>.
- (263) Cross, S.; Baroni, M.; Goracci, L.; Cruciani, G. GRID-Based Three-Dimensional Pharmacophores I: FLAPpharm, a Novel Approach for Pharmacophore Elucidation. *J. Chem. Inf. Model.* **2012**, *52* (10), 2587–2598. <https://doi.org/10.1021/ci300153d>.
- (264) Goodford, P. The Basic Principles of GRID. In *Molecular Interaction Fields*; John Wiley & Sons, Ltd, 2005; pp 1–25. <https://doi.org/10.1002/3527607676.ch1>.
- (265) Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. FLAP: GRID Molecular Interaction Fields in Virtual Screening. Validation Using the DUD Data Set. *J. Chem. Inf. Model.* **2010**, *50* (8), 1442–1450. <https://doi.org/10.1021/ci100221g>.
- (266) Grid-Derived Structure-Based 3D Pharmacophores and Their Performance Compared to Docking. *Drug Discov Today Technol* **2010**, *7* (4), e203-270. <https://doi.org/10.1016/j.ddtec.2010.09.002>.
- (267) Xia, J.; Tilahun, E. L.; Kebede, E. H.; Reid, T.-E.; Zhang, L.; Wang, X. S. Comparative Modeling and Benchmarking Data Sets for Human Histone Deacetylases and Sirtuin Families. *J. Chem. Inf. Model.* **2015**, *55* (2), 374–388. <https://doi.org/10.1021/ci5005515>.
- (268) Mysinger, M. M.; Carchia, M.; Irwin, John. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594. <https://doi.org/10.1021/jm300687e>.

- (269) *Specs - Compound Management services and Research Compounds for the Life Science industry*. <https://www.specs.net/index.php?page=2019041215290210#screeningcompounds> (accessed 2020-12-07).
- (270) Instant JChem 21.2.0., 2021.
- (271) RDKit: Open-Source Cheminformatics, 2021.
- (272) Vettigli, G. MiniSom 2.2.1., 2021. <https://github.com/JustGlowing/minisom>.
- (273) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res* **2019**, *47* (D1), D930–D940. <https://doi.org/10.1093/nar/gky1075>.
- (274) Chen, L.; Ai, T.; More, S. Therapeutic Compounds. US20160376238A1, December 29, 2016. <https://patents.google.com/patent/US20160376238A1/en?q=US+2016%2f0376238+A1> (accessed 2022-08-20).
- (275) Swain, M. C.; Cole, J. M. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. *J. Chem. Inf. Model.* **2016**, *56* (10), 1894–1904. <https://doi.org/10.1021/acs.jcim.6b00207>.
- (276) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON* 6.
- (277) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-Sampling Technique. *jair* **2002**, *16*, 321–357. <https://doi.org/10.1613/jair.953>.
- (278) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *Journal of Cheminformatics* **2018**, *10* (1), 4. <https://doi.org/10.1186/s13321-018-0258-y>.
- (279) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mane, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viegas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 19.
- (280) O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; others. Keras Tuner, 2019. <https://github.com/keras-team/keras-tuner>.
- (281) Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules* **2012**, *17* (5), 4791–4810. <https://doi.org/10.3390/molecules17054791>.
- (282) Chicco, D.; Jurman, G. An Invitation to Greater Use of Matthews Correlation Coefficient in Robotics and Artificial Intelligence. *Frontiers in Robotics and AI* **2022**, *9*.
- (283) Riniker, S.; Landrum, G. A. Similarity Maps - a Visualization Strategy for Molecular Fingerprints and Machine-Learning Methods. *Journal of Cheminformatics* **2013**, *5* (1), 43. <https://doi.org/10.1186/1758-2946-5-43>.
- (284) Marcotte, P. A.; Richardson, P. L.; Guo, J.; Barrett, L. W.; Xu, N.; Gunasekera, A.; Glaser, K. B. Fluorescence Assay of SIRT Protein Deacetylases Using an Acetylated Peptide Substrate and a Secondary Trypsin Reaction. *Anal Biochem* **2004**, *332* (1), 90–99. <https://doi.org/10.1016/j.ab.2004.05.039>.
- (285) Wang, P.; Chen, D.; An, J.; Lin, S.; Liu, T.; Li, Y.; Chen, L.; He, B. Development of a Single-Step Fluorogenic Sirtuin Assay and Its Applications for High-Throughput Screening. *Org. Biomol. Chem.* **2022**, *20* (6), 1243–1252. <https://doi.org/10.1039/D1OB02347K>.

- (286) Heger, V.; Tyni, J.; Hunyadi, A.; Horáková, L.; Lahtela-Kakkonen, M.; Rahnasto-Rilla, M. Quercetin Based Derivatives as Sirtuin Inhibitors. *Biomedicine & Pharmacotherapy* **2019**, *111*, 1326–1333. <https://doi.org/10.1016/j.biopha.2019.01.035>.
- (287) Kokkonen, P.; Rahnasto-Rilla, M.; Mellini, P.; Jarho, E.; Lahtela-Kakkonen, M.; Kokkola, T. Studying SIRT6 Regulation Using H3K56 Based Substrate and Small Molecules. *European Journal of Pharmaceutical Sciences* **2014**, *63*, 71–76. <https://doi.org/10.1016/j.ejps.2014.06.015>.
- (288) Mosmann, T. Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. *Journal of Immunological Methods* **1983**, *65* (1), 55–63. [https://doi.org/10.1016/0022-1759\(83\)90303-4](https://doi.org/10.1016/0022-1759(83)90303-4).
- (289) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748. <https://doi.org/10.1006/jmbi.1996.0897>.
- (290) Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein-Peptide Docking: Opportunities and Challenges. *Drug Discov Today* **2018**, *23* (8), 1530–1537. <https://doi.org/10.1016/j.drudis.2018.05.006>.
- (291) Hauser, A. S.; Windshügel, B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J. Chem. Inf. Model.* **2016**, *56* (1), 188–200. <https://doi.org/10.1021/acs.jcim.5b00234>.
- (292) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model* **2011**, *51* (1), 69–82. <https://doi.org/10.1021/ci100275a>.
- (293) Chen, F.; Liu, H.; Sun, H.; Pan, P.; Li, Y.; Li, D.; Hou, T. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 6. Capability to Predict Protein-Protein Binding Free Energies and Re-Rank Binding Poses Generated by Protein-Protein Docking. *Phys Chem Chem Phys* **2016**, *18* (32), 22129–22139. <https://doi.org/10.1039/c6cp03670h>.
- (294) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin Drug Discov* **2015**, *10* (5), 449–461. <https://doi.org/10.1517/17460441.2015.1032936>.
- (295) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>.
- (296) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (297) Vassetti, D.; Pagliai, M.; Procacci, P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J Chem Theory Comput* **2019**, *15* (3), 1983–1995. <https://doi.org/10.1021/acs.jctc.8b01039>.
- (298) Duchstein, P.; Neiss, C.; Görling, A.; Zahn, D. Molecular Mechanics Modeling of Azobenzene-Based Photoswitches. *J Mol Model* **2012**, *18* (6), 2479–2482. <https://doi.org/10.1007/s00894-011-1270-6>.
- (299) Woods, R. J.; Chappelle, R. Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates. *Theochem* **2000**, *527* (1–3), 149–156. [https://doi.org/10.1016/S0166-1280\(00\)00487-5](https://doi.org/10.1016/S0166-1280(00)00487-5).
- (300) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27* (11), 1575–1577. <https://doi.org/10.1093/bioinformatics/btr168>.

- (301) Biswas, M.; Burghardt, I. Azobenzene Photoisomerization-Induced Destabilization of B-DNA. *Biophys. J.* **2014**, *107* (4), 932–940. <https://doi.org/10.1016/j.bpj.2014.06.044>.
- (302) Yang, W.; Soares, J.; Greninger, P.; Edelman, E. J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J. A.; Thompson, I. R.; Ramaswamy, S.; Futreal, P. A.; Haber, D. A.; Stratton, M. R.; Benes, C.; McDermott, U.; Garnett, M. J. Genomics of Drug Sensitivity in Cancer (GDSC): A Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Research* **2013**, *41* (D1), D955–D961. <https://doi.org/10.1093/nar/gks1111>.
- (303) Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A. A.; Kim, S.; Wilson, C. J.; Lehár, J.; Kryukov, G. V.; Sonkin, D.; Reddy, A.; Liu, M.; Murray, L.; Berger, M. F.; Monahan, J. E.; Morais, P.; Meltzer, J.; Korejwa, A.; Jané-Valbuena, J.; Mapa, F. A.; Thibault, J.; Bric-Furlong, E.; Raman, P.; Shipway, A.; Engels, I. H.; Cheng, J.; Yu, G. K.; Yu, J.; Aspesi, P.; de Silva, M.; Jagtap, K.; Jones, M. D.; Wang, L.; Hatton, C.; Palesscandolo, E.; Gupta, S.; Mahan, S.; Sougnez, C.; Onofrio, R. C.; Liefeld, T.; MacConaill, L.; Winckler, W.; Reich, M.; Li, N.; Mesirov, J. P.; Gabriel, S. B.; Getz, G.; Ardlie, K.; Chan, V.; Myer, V. E.; Weber, B. L.; Porter, J.; Warmuth, M.; Finan, P.; Harris, J. L.; Meyerson, M.; Golub, T. R.; Morrissey, M. P.; Sellers, W. R.; Schlegel, R.; Garraway, L. A. The Cancer Cell Line Encyclopedia Enables Predictive Modeling of Anticancer Drug Sensitivity. *Nature* **2012**, *483* (7391), 603–607. <https://doi.org/10.1038/nature11003>.
- (304) Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update). *Nucleic Acids Research* **2019**, *47* (W1), W191–W198. <https://doi.org/10.1093/nar/gkz369>.
- (305) Chou, T. C.; Talalay, P. Quantitative Analysis of Dose-Effect Relationships: The Combined Effects of Multiple Drugs or Enzyme Inhibitors. *Adv Enzyme Regul* **1984**, *22*, 27–55. [https://doi.org/10.1016/0065-2571\(84\)90007-4](https://doi.org/10.1016/0065-2571(84)90007-4).
- (306) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9* (4), 2000–2009. <https://doi.org/10.1021/ct300878a>.
- (307) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1), 015102. <https://doi.org/10.1063/1.4811489>.
- (308) Sittel, F.; Stock, G. Perspective: Identification of Collective Variables and Metastable States of Protein Dynamics. *J. Chem. Phys.* **2018**, *149* (15), 150901. <https://doi.org/10.1063/1.5049637>.
- (309) Benabderrahmane, M.; Bureau, R.; Voisin-Chiret, A. S.; Santos, J. S. O. Cryptic Pockets Repository through Pocket Dynamics Tracking and Metadynamics on Essential Dynamics Space: Applications to Mcl-1. *J. Chem. Inf. Model.* **2021**, *61* (11), 5581–5588. <https://doi.org/10.1021/acs.jcim.1c00660>.
- (310) Sicard, F.; Senet, P. Reconstructing the Free-Energy Landscape of Met-Enkephalin Using Dihedral Principal Component Analysis and Well-Tempered Metadynamics. *J. Chem. Phys.* **2013**, *138* (23), 235101. <https://doi.org/10.1063/1.4810884>.
- (311) Maurer, B.; Rumpf, T.; Scharfe, M.; Stolfa, D. A.; Schmitt, M. L.; He, W.; Verdin, E.; Sippl, W.; Jung, M. Inhibitors of the NAD⁺-Dependent Protein Desuccinylase and Demalonylase Sirt5. *ACS Med. Chem. Lett.* **2012**, *3* (12), 1050–1053. <https://doi.org/10.1021/ml3002709>.
- (312) Dai, X.; Cheng, H.; Bai, Z.; Li, J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J. Cancer* **2017**, *8* (16), 3131–3141. <https://doi.org/10.7150/jca.18457>.
- (313) Järvenpää, J.; Rahnasto-Rilla, M.; Lahtela-Kakkonen, M.; Küblbeck, J. Profiling the Regulatory Interplay of BET Bromodomains and Sirtuins in Cancer Cell Lines. *Biomedicine & Pharmacotherapy* **2022**, *147*, 112652. <https://doi.org/10.1016/j.biopha.2022.112652>.

- (314) Pastushenko, I.; Blanpain, C. EMT Transition States during Tumor Progression and Metastasis. *Trends in Cell Biology* **2019**, *29* (3), 212–226. <https://doi.org/10.1016/j.tcb.2018.12.001>.
- (315) Dongre, A.; Weinberg, R. A. New Insights into the Mechanisms of Epithelial–Mesenchymal Transition and Implications for Cancer. *Nat Rev Mol Cell Biol* **2019**, *20* (2), 69–84. <https://doi.org/10.1038/s41580-018-0080-4>.
- (316) Chen, J.; Chan, A. W. H.; To, K.-F.; Chen, W.; Zhang, Z.; Ren, J.; Song, C.; Cheung, Y.-S.; Lai, P. B. S.; Cheng, S.-H.; Ng, M. H. L.; Huang, A.; Ko, B. C. B. SIRT2 Overexpression in Hepatocellular Carcinoma Mediates Epithelial to Mesenchymal Transition by Protein Kinase B/Glycogen Synthase Kinase-3 β / β -Catenin Signaling. *Hepatology* **2013**, *57* (6), 2287–2298. <https://doi.org/10.1002/hep.26278>.
- (317) Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of Mixed IC50 Data – A Statistical Analysis. *PLoS One* **2013**, *8* (4), e61007. <https://doi.org/10.1371/journal.pone.0061007>.
- (318) Suvannang, N.; Preeyanon, L.; Ahmad Malik, A.; Schaduangrat, N.; Shoombuatong, W.; Worachartcheewan, A.; Tantimongcolwat, T.; Nantasenamat, C. Probing the Origin of Estrogen Receptor Alpha Inhibition via Large-Scale QSAR Study. *RSC Advances* **2018**, *8* (21), 11344–11356. <https://doi.org/10.1039/C7RA10979B>.
- (319) Zakharov, A. V.; Zhao, T.; Nguyen, D.-T.; Peryea, T.; Sheils, T.; Yasgar, A.; Huang, R.; Southall, N.; Simeonov, A. Novel Consensus Architecture To Improve Performance of Large-Scale Multitask Deep Learning QSAR Models. *J. Chem. Inf. Model.* **2019**, *59* (11), 4613–4624. <https://doi.org/10.1021/acs.jcim.9b00526>.
- (320) Li, S.; Ding, Y.; Chen, M.; Chen, Y.; Kirchmair, J.; Zhu, Z.; Wu, S.; Xia, J. HDAC3i-Finder: A Machine Learning-Based Computational Tool to Screen for HDAC3 Inhibitors. *Mol Inform* **2021**, *40* (3), e2000105. <https://doi.org/10.1002/minf.202000105>.
- (321) Li, R.; Tian, Y.; Yang, Z.; Ji, Y.; Ding, J.; Yan, A. Classification Models and SAR Analysis on HDAC1 Inhibitors Using Machine Learning Methods. *Mol Divers* **2022**. <https://doi.org/10.1007/s11030-022-10466-w>.
- (322) Machado, L. A.; Krempser, E.; Guimarães, A. C. R. A Machine Learning-Based Virtual Screening for Natural Compounds Capable of Inhibiting the HIV-1 Integrase. *Front. Drug. Discov.* **2022**, *2*. <https://doi.org/10.3389/fddsv.2022.954911>.
- (323) Tropsha, A.; Golbraikh, A. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Curr Pharm Des* **2007**, *13* (34), 3494–3504. <https://doi.org/10.2174/138161207782794257>.
- (324) Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection. *J. Chem. Inf. Model.* **2012**, *52* (8), 2044–2058. <https://doi.org/10.1021/ci300084j>.
- (325) Blum, C. A.; Ellis, J. L.; Loh, C.; Ng, P. Y.; Perni, R. B.; Stein, R. L. SIRT1 Modulation as a Novel Approach to the Treatment of Diseases of Aging. *J. Med. Chem.* **2011**, *54* (2), 417–432. <https://doi.org/10.1021/jm100861p>.
- (326) Karatas, H.; Townsend, E. C.; Bernard, D.; Dou, Y.; Wang, S. Analysis of the Binding of Mixed Lineage Leukemia 1 (MLL1) and Histone 3 Peptides to WD Repeat Domain 5 (WDR5) for the Design of Inhibitors of the MLL1–WDR5 Interaction. *J Med Chem* **2010**, *53* (14), 5179–5185. <https://doi.org/10.1021/jm100139b>.
- (327) Liao, J.-M.; Wang, Y.-T.; Lin, C.-L. S. A Fragment-Based Docking Simulation for Investigating Peptide-Protein Bindings. *Phys Chem Chem Phys* **2017**, *19* (16), 10436–10442. <https://doi.org/10.1039/c6cp07136h>.
- (328) Ruthenburg, A. J.; Wang, W.; Graybosch, D. M.; Li, H.; Allis, C. D.; Patel, D. J.; Verdine, G. L. Histone H3 Recognition and Presentation by the WDR5 Module of the MLL1 Complex. *Nat Struct Mol Biol* **2006**, *13* (8), 704–712. <https://doi.org/10.1038/nsmb1119>.

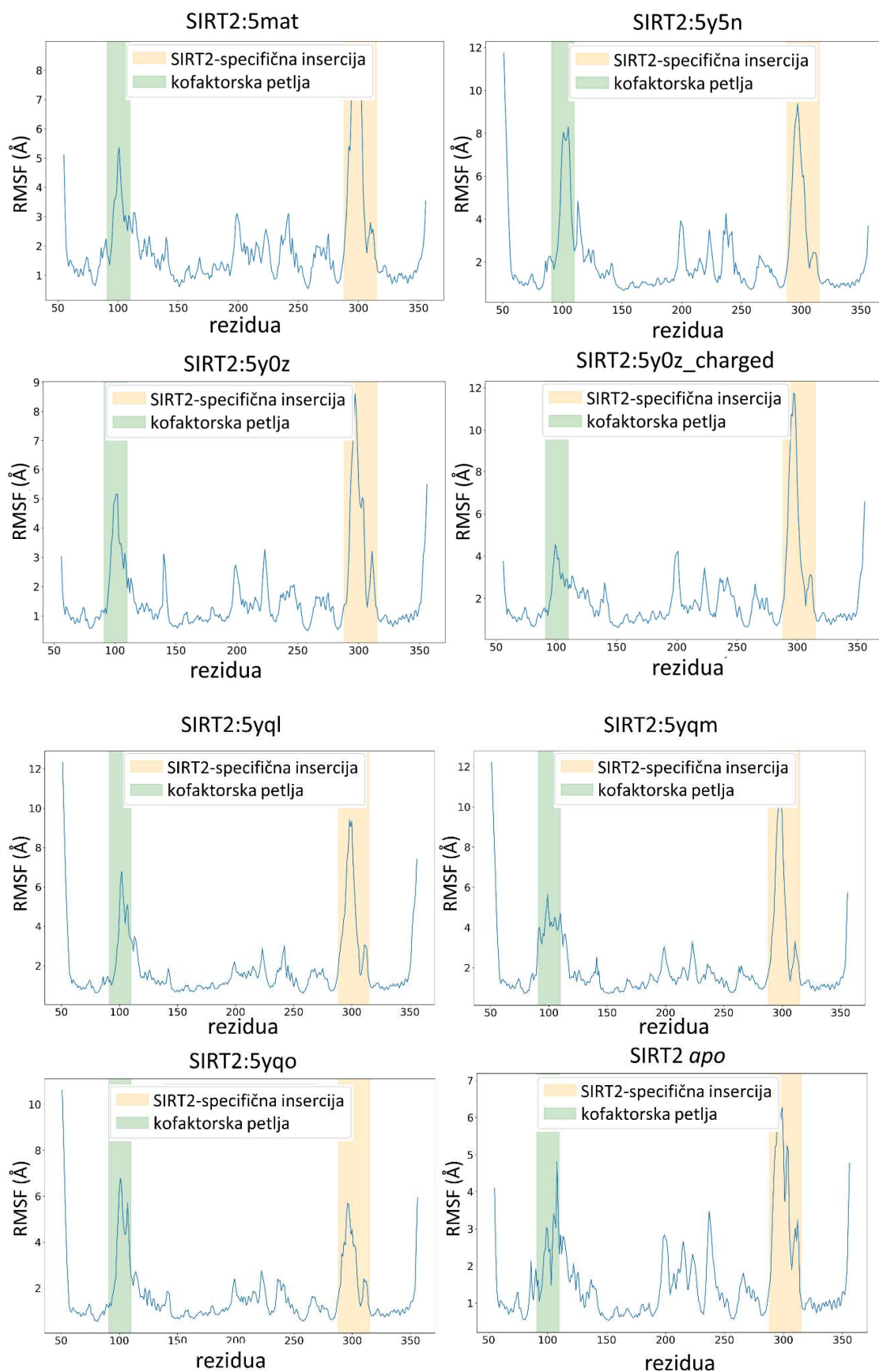
- (329) Qvit, N.; Rubin, S. J. S.; Urban, T. J.; Mochly-Rosen, D.; Gross, E. R. Peptidomimetic Therapeutics: Scientific Approaches and Opportunities. *Drug Discov Today* **2017**, *22* (2), 454–462. <https://doi.org/10.1016/j.drudis.2016.11.003>.
- (330) Atilgan, C.; Atilgan, A. R. Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein. *PLOS Computational Biology* **2009**, *5* (10), e1000544. <https://doi.org/10.1371/journal.pcbi.1000544>.
- (331) Dou, Y.; Milne, T. A.; Ruthenburg, A. J.; Lee, S.; Lee, J. W.; Verdine, G. L.; Allis, C. D.; Roeder, R. G. Regulation of MLL1 H3K4 Methyltransferase Activity by Its Core Components. *Nature Structural & Molecular Biology* **2006**, *13* (8), 713–719. <https://doi.org/10.1038/nsmb1128>.
- (332) McGuigan, A.; Kelly, P.; Turkington, R. C.; Jones, C.; Coleman, H. G.; McCain, R. S. Pancreatic Cancer: A Review of Clinical Diagnosis, Epidemiology, Treatment and Outcomes. *World J Gastroenterol* **2018**, *24* (43), 4846–4861. <https://doi.org/10.3748/wjg.v24.i43.4846>.
- (333) Koltai, T.; Reshkin, S. J.; Carvalho, T. M. A.; Di Molfetta, D.; Greco, M. R.; Alfarouk, K. O.; Cardone, R. A. Resistance to Gemcitabine in Pancreatic Ductal Adenocarcinoma: A Physiopathologic and Pharmacologic Review. *Cancers* **2022**, *14* (10), 2486. <https://doi.org/10.3390/cancers14102486>.
- (334) Rahib, L.; Wehner, M. R.; Matrisian, L. M.; Nead, K. T. Estimated Projection of US Cancer Incidence and Death to 2040. *JAMA Netw Open* **2021**, *4* (4), e214708. <https://doi.org/10.1001/jamanetworkopen.2021.4708>.
- (335) Labori, K. J.; Katz, M. H.; Tzeng, C. W.; Bjørneth, B. A.; Cvancarova, M.; Edwin, B.; Kure, E. H.; Eide, T. J.; Dueland, S.; Buanes, T.; Gladhaug, I. P. Impact of Early Disease Progression and Surgical Complications on Adjuvant Chemotherapy Completion Rates and Survival in Patients Undergoing the Surgery First Approach for Resectable Pancreatic Ductal Adenocarcinoma - A Population-Based Cohort Study. *Acta Oncol* **2016**, *55* (3), 265–277. <https://doi.org/10.3109/0284186X.2015.1068445>.
- (336) Orth, M.; Metzger, P.; Gerum, S.; Mayerle, J.; Schneider, G.; Belka, C.; Schnurr, M.; Lauber, K. Pancreatic Ductal Adenocarcinoma: Biological Hallmarks, Current Status, and Future Perspectives of Combined Modality Treatment Approaches. *Radiation Oncology* **2019**, *14* (1), 141. <https://doi.org/10.1186/s13014-019-1345-6>.
- (337) Yu, S.; Zhang, C.; Xie, K.-P. Therapeutic Resistance of Pancreatic Cancer: Roadmap to Its Reversal. *Biochim Biophys Acta Rev Cancer* **2021**, *1875* (1), 188461. <https://doi.org/10.1016/j.bbcan.2020.188461>.
- (338) Xiang, X.-S.; Li, P.-C.; Wang, W.-Q.; Liu, L. Histone Deacetylases: A Novel Class of Therapeutic Targets for Pancreatic Cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2022**, *1877* (1), 188676. <https://doi.org/10.1016/j.bbcan.2022.188676>.
- (339) Qiu, T.; Zhou, L.; Zhu, W.; Wang, T.; Wang, J.; Shu, Y.; Liu, P. Effects of Treatment with Histone Deacetylase Inhibitors in Solid Tumors: A Review Based on 30 Clinical Trials. *Future Oncology* **2013**, *9* (2), 255–269. <https://doi.org/10.2217/fon.12.173>.
- (340) Laschanzky, R. S.; Humphrey, L. E.; Ma, J.; Smith, L. M.; Enke, T. J.; Shukla, S. K.; Dasgupta, A.; Singh, P. K.; Howell, G. M.; Brattain, M. G.; Ly, Q. P.; Black, A. R.; Black, J. D. Selective Inhibition of Histone Deacetylases 1/2/6 in Combination with Gemcitabine: A Promising Combination for Pancreatic Cancer Therapy. *Cancers* **2019**, *11* (9), 1327. <https://doi.org/10.3390/cancers11091327>.
- (341) Djokovic, N.; Djuric, A.; Ruzic, D.; Srdic-Rajic, T.; Nikolic, K. Correlating Basal Gene Expression across Chemical Sensitivity Data to Screen for Novel Synergistic Interactors of HDAC Inhibitors in Pancreatic Carcinoma. *Pharmaceuticals* **2023**, *16* (2), 294. <https://doi.org/10.3390/ph16020294>.
- (342) Susanto, J. M.; Colvin, E. K.; Pinese, M.; Chang, D. K.; Pajic, M.; Mawson, A.; Caldon, C. E.; Musgrove, E. A.; Henshall, S. M.; Sutherland, R. L.; Biankin, A. V.; Scarlett, C. J. The Epigenetic Agents Suberoylanilide Hydroxamic Acid and 5-AZA-2' Deoxycytidine Decrease

- Cell Proliferation, Induce Cell Death and Delay the Growth of MiaPaCa2 Pancreatic Cancer Cells in Vivo. *International Journal of Oncology* **2015**, *46* (5), 2223–2230. <https://doi.org/10.3892/ijo.2015.2894>.
- (343) Poklepovic, A. S.; Fields, E. C.; Bandyopadhyay, D.; Tombes, M. B.; Kmiecik, M.; McGuire, W. P.; Gordon, S. W.; Kaplan, B. J.; Myers, J. L.; Matin, K.; Patel, B.; Kothadia, S.; Dent, P. A Phase 1 Study of Neoadjuvant Chemotherapy Followed by Concurrent Chemoradiation with Gemcitabine, Sorafenib, and Vorinostat in Pancreatic Cancer. *JCO* **2021**, *39* (15_suppl), e16268–e16268. https://doi.org/10.1200/JCO.2021.39.15_suppl.e16268.
- (344) Fu, X.; Zhang, X.; Yang, H.; Xu, X.; Hu, Z.; Yan, J.; Zheng, X.; Wei, R.; Zhang, Z.; Tang, S.; Geng, M.; Huang, X. CUDC-907 Displays Potent Antitumor Activity against Human Pancreatic Adenocarcinoma in Vitro and in Vivo through Inhibition of HDAC6 to Downregulate c-Myc Expression. *Acta Pharmacol Sin* **2019**, *40* (5), 677–688. <https://doi.org/10.1038/s41401-018-0108-5>.
- (345) Biermann, M.; Quintero, C.; Ferguson, P.; Rajbhandari, N.; Park, D. E.; Patel, H.; Reya, T. Repurposing HDAC and MTOR Inhibitors for Pancreatic Cancer. *JCO* **2022**, *40* (16_suppl), e16234–e16234. https://doi.org/10.1200/JCO.2022.40.16_suppl.e16234.
- (346) Zhang, X.; Zegar, T.; Weiser, T.; Hamdan, F. H.; Berger, B.-T.; Lucas, R.; Balourdas, D.-I.; Ladigan, S.; Cheung, P. F.; Liffers, S.-T.; Trajkovic-Arsic, M.; Scheffler, B.; Joerger, A. C.; Hahn, S. A.; Johnsen, S. A.; Knapp, S.; Siveke, J. T. Characterization of a Dual BET/HDAC Inhibitor for Treatment of Pancreatic Ductal Adenocarcinoma. *International Journal of Cancer* **2020**, *147* (10), 2847–2861. <https://doi.org/10.1002/ijc.33137>.
- (347) Kretz, A.-L.; Schaum, M.; Richter, J.; Kitzig, E. F.; Engler, C. C.; Leithäuser, F.; Henne-Bruns, D.; Knippschild, U.; Lemke, J. CDK9 Is a Prognostic Marker and Therapeutic Target in Pancreatic Cancer. *Tumour Biol* **2017**, *39* (2), 1010428317694304. <https://doi.org/10.1177/1010428317694304>.
- (348) Iriana, S.; Ahmed, S.; Gong, J.; Annamalai, A. A.; Tuli, R.; Hendifar, A. E. Targeting MTOR in Pancreatic Ductal Adenocarcinoma. *Front Oncol* **2016**, *6*, 99. <https://doi.org/10.3389/fonc.2016.00099>.
- (349) Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res* **2016**, *44* (D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
- (350) KEGG PATHWAY: *Sphingolipid signaling pathway - Homo sapiens (human)*. <https://www.genome.jp/pathway/hsa04071> (accessed 2022-10-09).
- (351) Slenter, D. N.; Kutmon, M.; Hanspers, K.; Riutta, A.; Windsor, J.; Nunes, N.; Mélius, J.; Cirillo, E.; Coort, S. L.; Digles, D.; Ehrhart, F.; Giesbertz, P.; Kalafati, M.; Martens, M.; Miller, R.; Nishida, K.; Rieswijk, L.; Waagmeester, A.; Eijssen, L. M. T.; Evelo, C. T.; Pico, A. R.; Willighagen, E. L. WikiPathways: A Multifaceted Pathway Database Bridging Metabolomics to Other Omics Research. *Nucleic Acids Res* **2018**, *46* (D1), D661–D667. <https://doi.org/10.1093/nar/gkx1064>.
- (352) Ruzic, D.; Ellinger, B.; Djokovic, N.; Santibanez, J. F.; Gul, S.; Beljkas, M.; Djuric, A.; Ganesan, A.; Pavic, A.; Srdic-Rajic, T.; Petkovic, M.; Nikolic, K. Discovery of 1-Benzhydryl-Piperazine-Based HDAC Inhibitors with Anti-Breast Cancer Activity: Synthesis, Molecular Modeling, In Vitro and In Vivo Biological Evaluation. *Pharmaceutics* **2022**, *14* (12), 2600. <https://doi.org/10.3390/pharmaceutics14122600>.
- (353) Vennin, C.; Rath, N.; Pajic, M.; Olson, M. F.; Timpson, P. Targeting ROCK Activity to Disrupt and Prime Pancreatic Cancer for Chemotherapy. *Small GTPases* **2020**, *11* (1), 45–52. <https://doi.org/10.1080/21541248.2017.1345712>.
- (354) Rath, N.; Munro, J.; Cutiongco, M. F.; Jagiełło, A.; Gadegaard, N.; McGarry, L.; Unbekandt, M.; Michalopoulou, E.; Kamphorst, J. J.; Sumpton, D.; Mackay, G.; Vennin, C.; Pajic, M.; Timpson, P.; Olson, M. F. Rho Kinase Inhibition by AT13148 Blocks Pancreatic Ductal Adenocarcinoma Invasion and Tumor Growth. *Cancer Research* **2018**, *78* (12), 3321–3336. <https://doi.org/10.1158/0008-5472.CAN-17-1339>.

- (355) Lepley, D.; Paik, J.-H.; Hla, T.; Ferrer, F. The G Protein-Coupled Receptor S1P2 Regulates Rho/Rho Kinase Pathway to Inhibit Tumor Cell Migration. *Cancer Res* **2005**, *65* (9), 3788–3795. <https://doi.org/10.1158/0008-5472.CAN-04-2311>.
- (356) Hait, N. C.; Wise, L. E.; Allegood, J. C.; O'Brien, M.; Avni, D.; Reeves, T. M.; Knapp, P. E.; Lu, J.; Luo, C.; Miles, M. F.; Milstien, S.; Lichtman, A. H.; Spiegel, S. Active, Phosphorylated Fingolimod Inhibits Histone Deacetylases and Facilitates Fear Extinction Memory. *Nat Neurosci* **2014**, *17* (7), 971–980. <https://doi.org/10.1038/nn.3728>.
- (357) Sriram, K.; Salmerón, C.; Wiley, S. Z.; Insel, P. A. GPCRs in Pancreatic Adenocarcinoma: Contributors to Tumour Biology and Novel Therapeutic Targets. *Br J Pharmacol* **2020**, *177* (11), 2434–2455. <https://doi.org/10.1111/bph.15028>.
- (358) Huwiler, A.; Zangemeister-Wittke, U. The Sphingosine 1-Phosphate Receptor Modulator Fingolimod as a Therapeutic Agent: Recent Findings and New Perspectives. *Pharmacol Ther* **2018**, *185*, 34–49. <https://doi.org/10.1016/j.pharmthera.2017.11.001>.

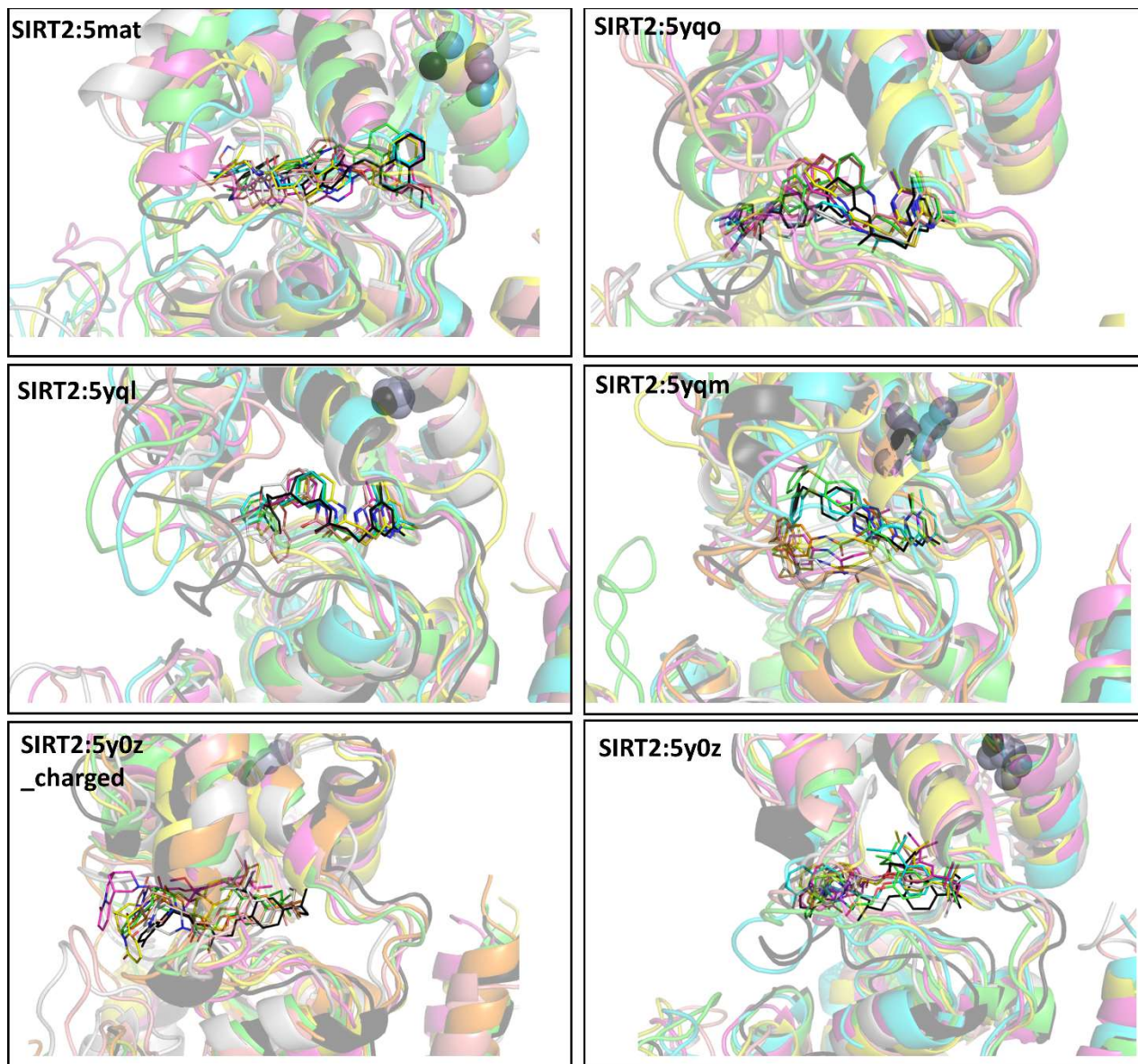
7. Prilozi

Prilog 7.1. RMSF analiza trajektorija iz SIRT2 MD simulacija.



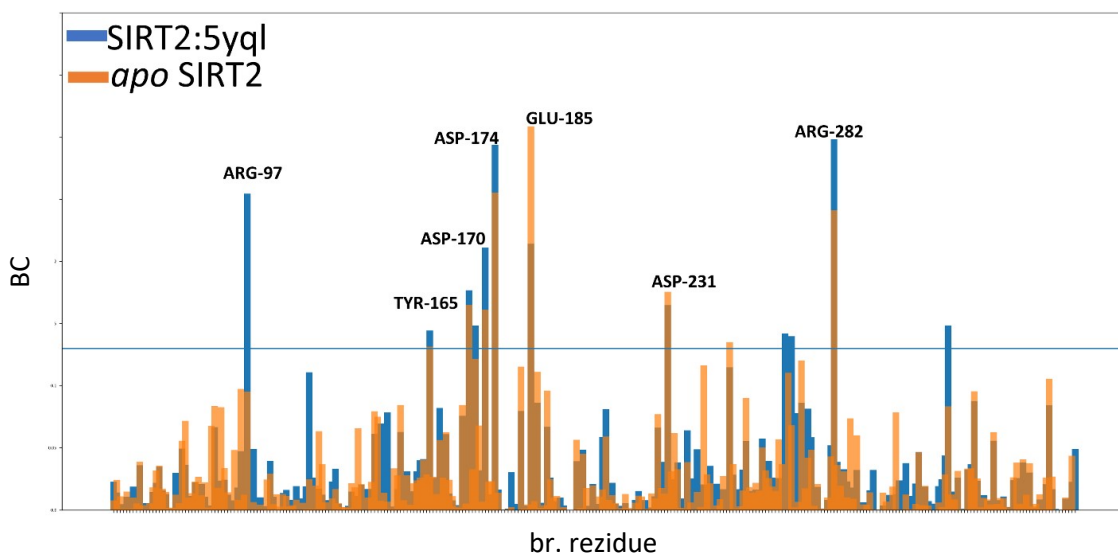
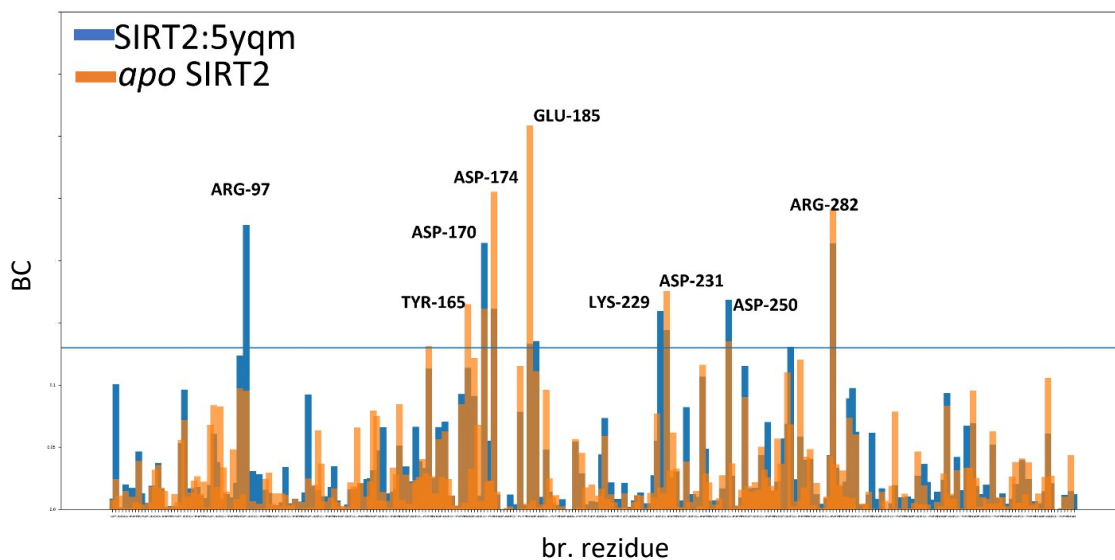
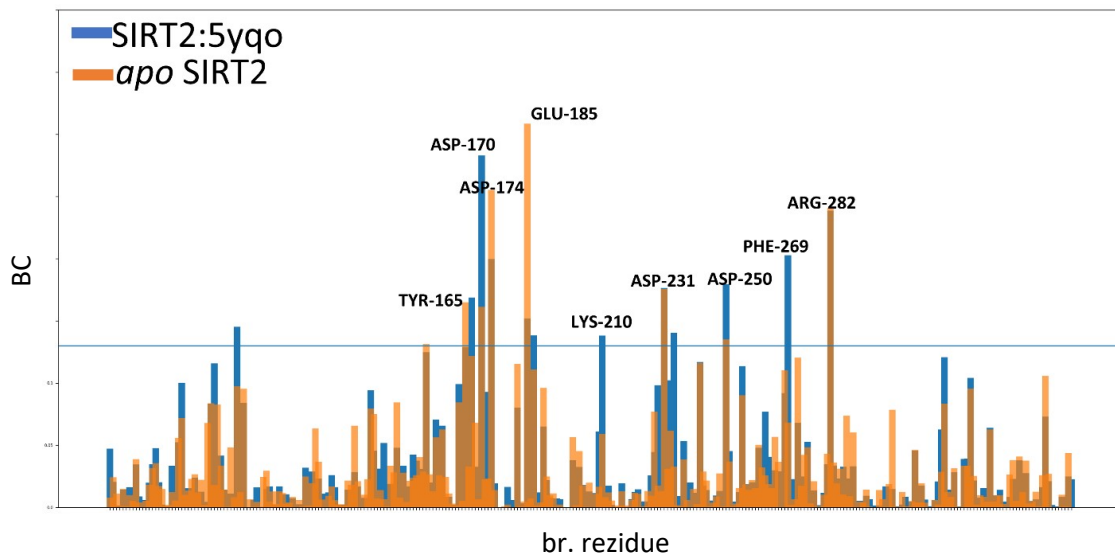
Slika P1. RMSF analiza trajektorija iz SIRT2 MD simulacija. Žutom i zelenom bojom su naznačeni specifični delovi strukture SIRT2 proteina.

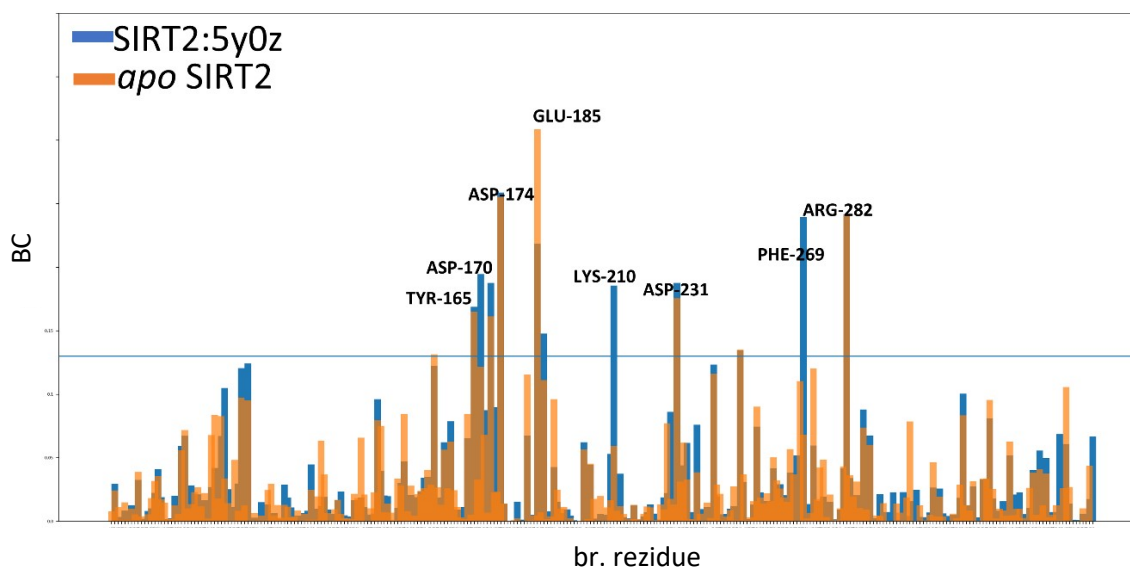
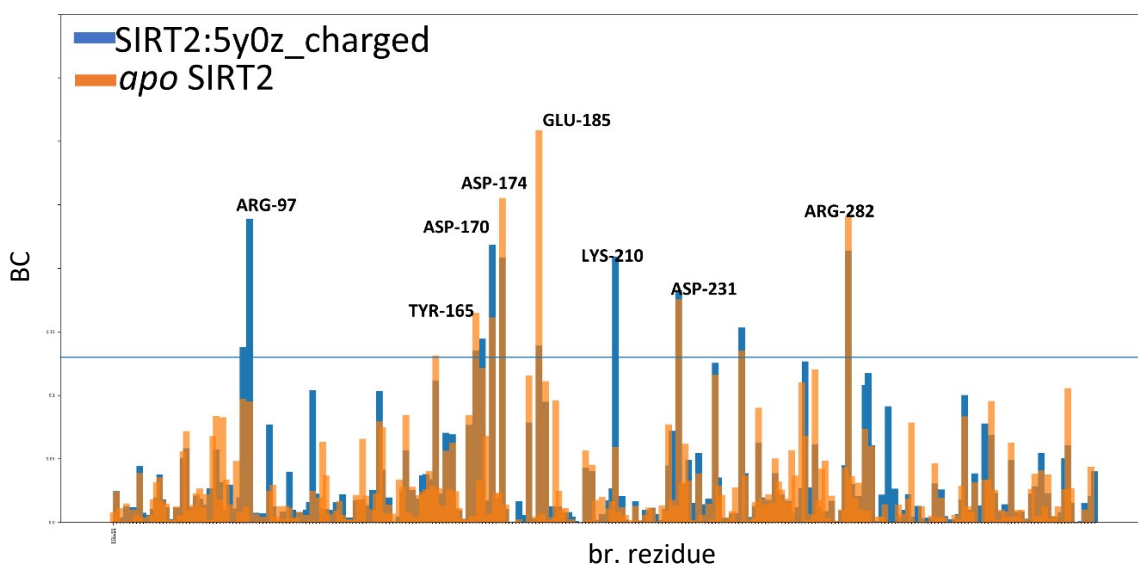
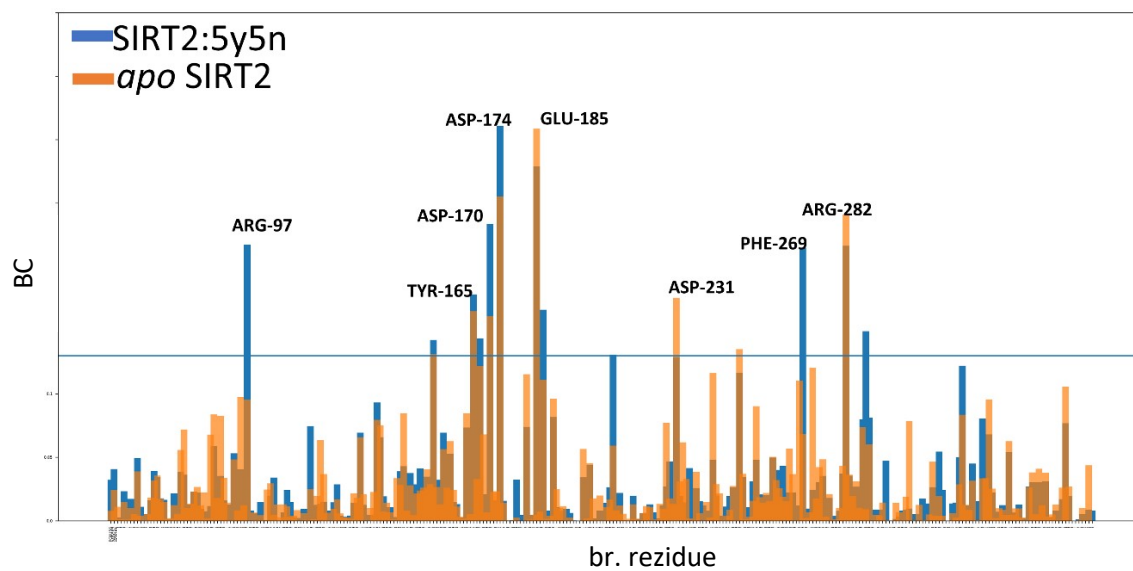
Prilog 7.2. Alternativni vezivni modovi liganada (centroidi trajektorija) dobijeni nakon klasičnih MD simulacija SIRT2 sistema.

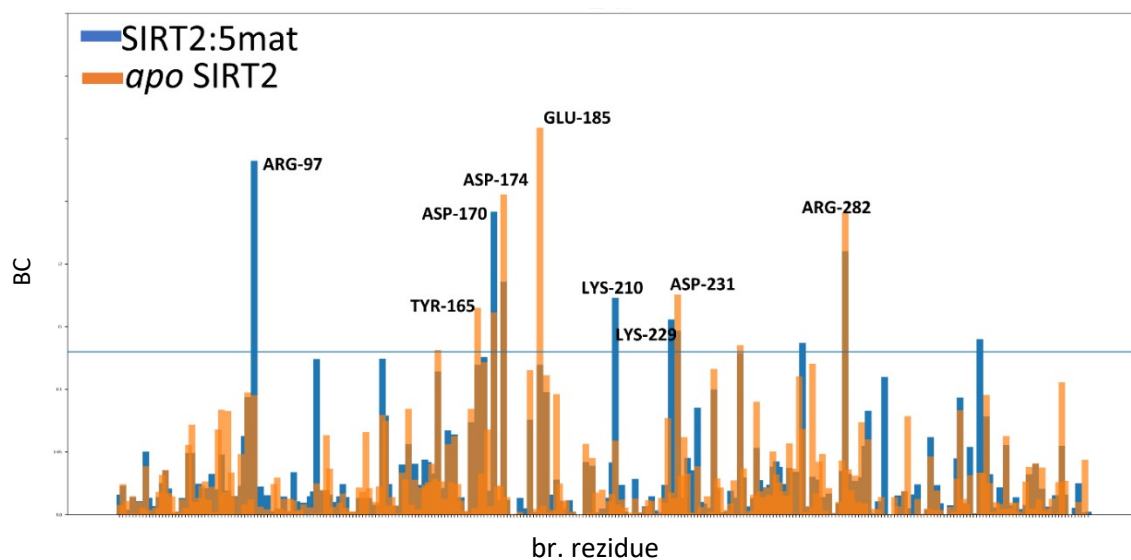


Slika P2. Alternativni vezivni modovi liganada (centroidi trajektorija) dobijeni nakon klasičnih MD simulacija SIRT2 sistema. Konformacije liganada koje odgovaraju kristalografskim strukturama prikazane su crnom bojom.

Prilog 7.3. Rezultati BC analize dobijeni analizom MD simulaicija SIRT2.

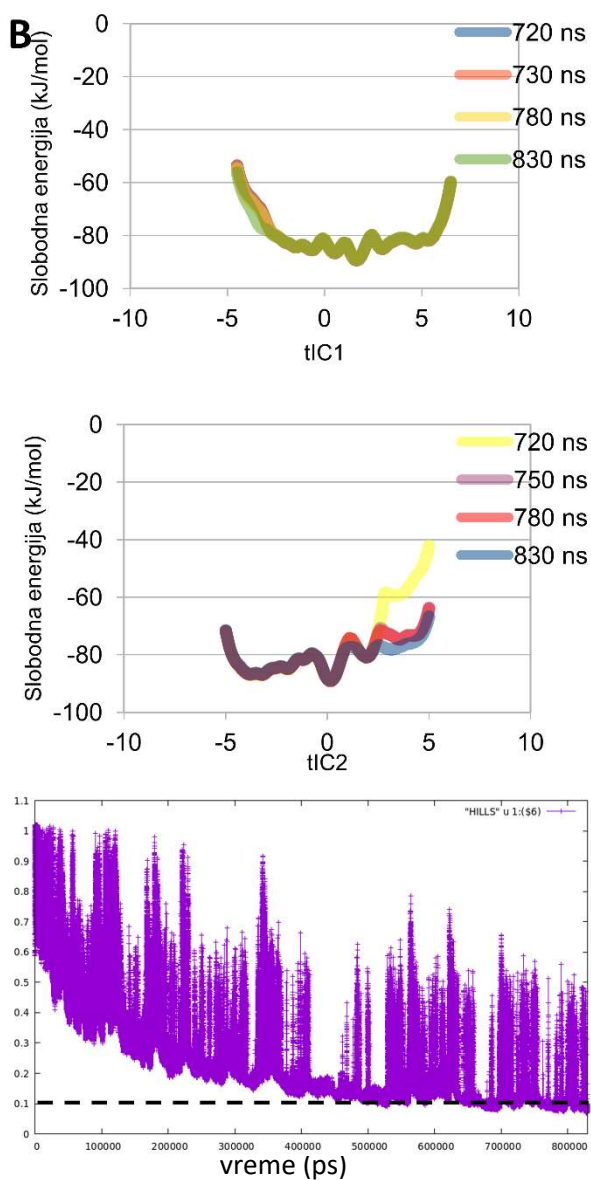
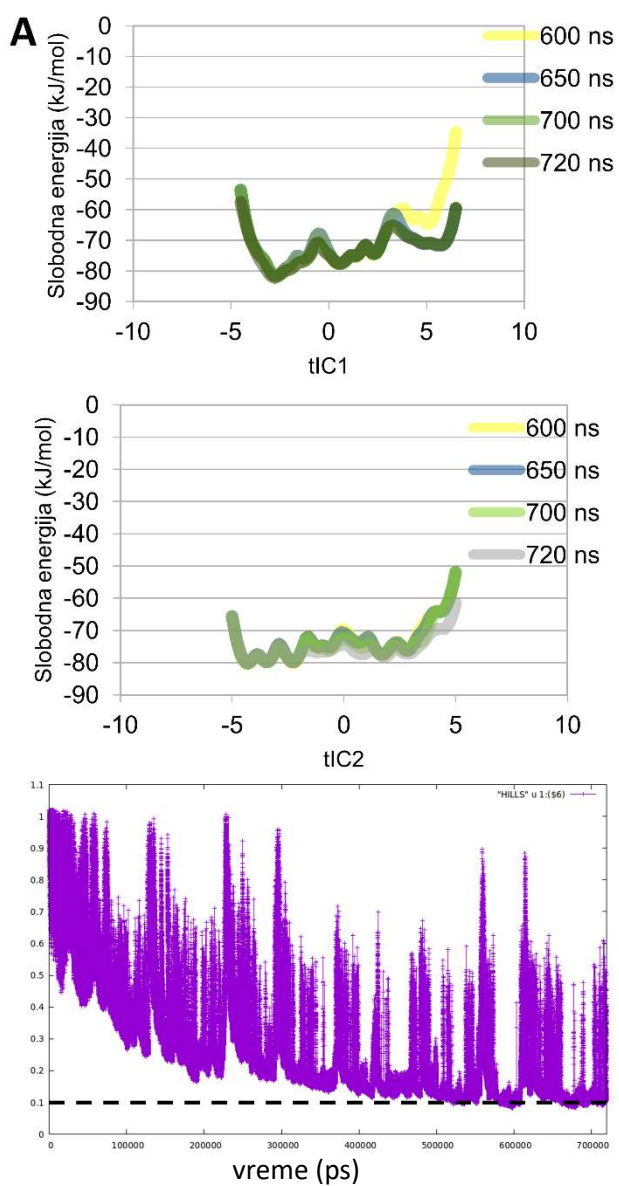


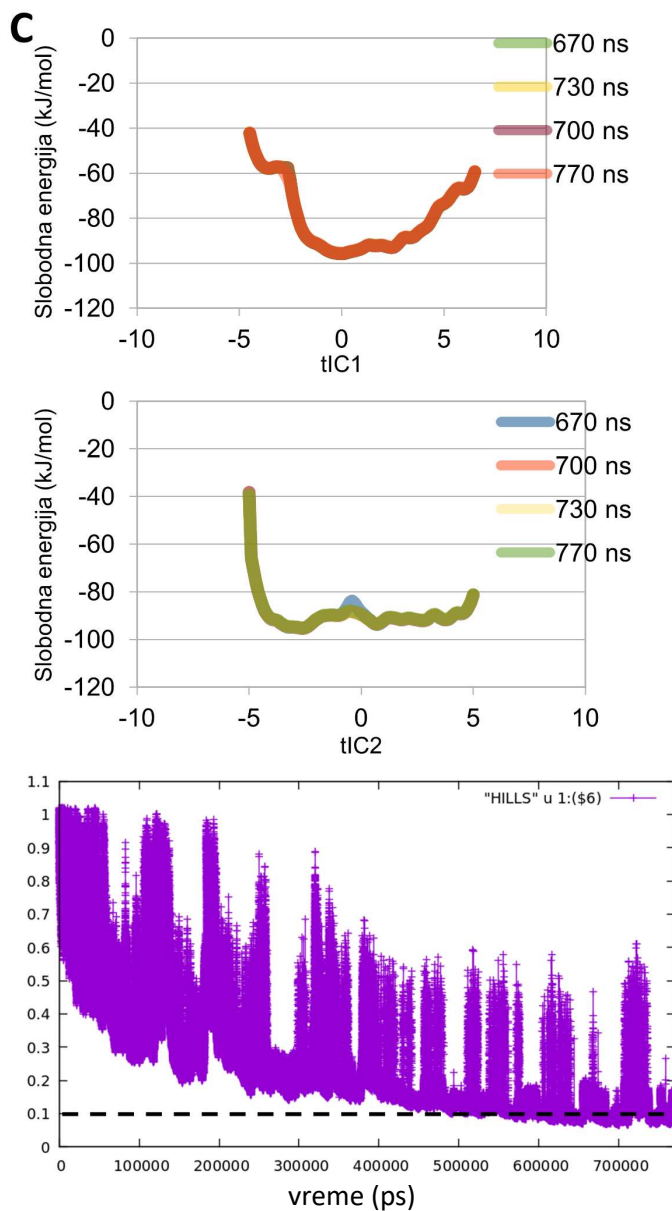




Slika P3. Rezultati BC analize dobijeni analizom *apo/olo* MD simulacija SIRT2 za svaki od simuliranih sistema. Rezidue sa najizraženijim BC vrednostima su obeležene slovima.

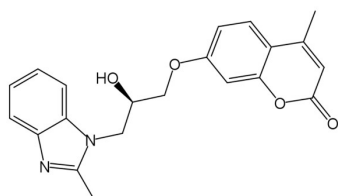
Prilog 7.4. Prikaz konvergencije metadinamičkih simulacija SIRT2 sistema.



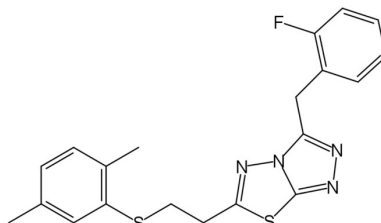


Slika P4. Prikaz konvergencije metadinamičkih simulacija SIRT2 sistema. A – SIRT2:5yqo; B – SIRT2:5y5n; C – SIRT2:apo. Isprekidana linija ukazuje na kvalitativni kriterijum konvergencije razmatran kao pad visine *Gaussian*-a inicijalnog potencijala pristrasnosti na ispod 10% od početnog.

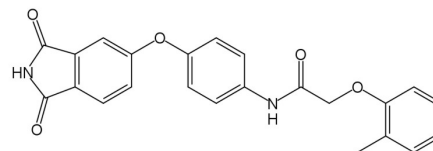
Prilog 7.5. Supstance testirane nakon prospektivnog SBVS.



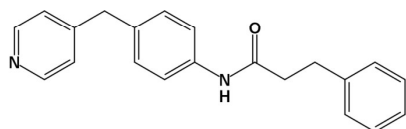
ND4
 SPECS ID: AP-124/43383635
 %Inh @200 μ M (SIRT1) = 11.83
 %Inh @200 μ M (SIRT2) = 31.85
 %Inh @200 μ M (SIRT6) = 4.2



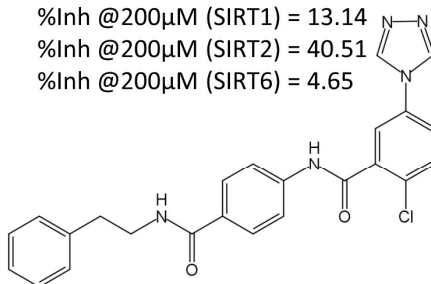
ND10
 SPECS ID: AP-501/43286878
 %Inh @200 μ M (SIRT1) = 13.14
 %Inh @200 μ M (SIRT2) = 40.51
 %Inh @200 μ M (SIRT6) = 4.65



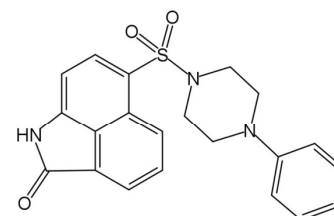
ND11
 SPECS ID: AN-979/41713669
 %Inh @200 μ M (SIRT1) = 10.70
 %Inh @200 μ M (SIRT2) = 55.53
 %Inh @200 μ M (SIRT6) = 16.88



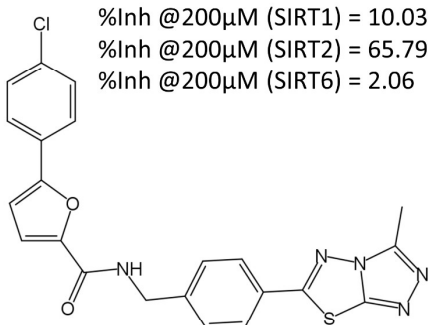
ND18
 SPECS ID: AO-080/40098868
 %Inh @200 μ M (SIRT1) = 10.03
 %Inh @200 μ M (SIRT2) = 65.79
 %Inh @200 μ M (SIRT6) = 2.06



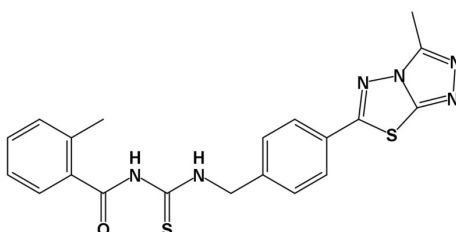
ND20
 SPECS ID: AN-465/42834327
 %Inh @200 μ M (SIRT1) = 10.41
 %Inh @200 μ M (SIRT2) = 45.07
 %Inh @200 μ M (SIRT6) = 5.57



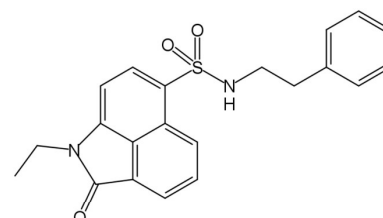
ND26
 SPECS ID: AQ-390/43238263
 %Inh @200 μ M (SIRT1) = 16.14
 %Inh @200 μ M (SIRT2) = 30.01
 %Inh @200 μ M (SIRT6) = 5.54



ND57
 SPECS ID: AP-970/43202690
 %Inh @200 μ M (SIRT1) = 15.24
 %Inh @200 μ M (SIRT2) = 47.22
 %Inh @200 μ M (SIRT6) = 4.8



ND85
 SPECS ID: AP-970/43202520
 %Inh @200 μ M (SIRT1) = 2.66
 %Inh @200 μ M (SIRT2) = 65.70
 %Inh @200 μ M (SIRT6) = 9.99



ND134
 SPECS ID: AQ-390/42869130
 %Inh @200 μ M (SIRT1) = 55.20
 %Inh @200 μ M (SIRT2) = 70.17
 %Inh @200 μ M (SIRT6) = N.D.

Slika P5. Strukture i *in vitro* rezultati testiranih inhibitora nakon prospektivnog SBVS SIRT2.

Prilog 7.6. Poređenje različitih SBVS modela.

Tabela P1. Rangiranje devet testiranih jedinjenja primenom različitih SBVS modela u prospektivnom skriningu SPECS baze.

Ligand	Kristalografske strukture						Finalni multistrukturalni model
	SIRT2:5mat	SIRT2:5y5n	SIRT2:5yqo	SIRT2:5yql	SIRT2:5yqm	SIRT2:5y0z	
ND26	40981	7180	55400	3741	10008	2321	67 ^t
ND4	82344	48521	46305	8240	4041	38405	118
ND10	41183	21162	83667	63985	121825	77029	441
ND20	49042	57481	25623	7768	22286	14920	89
ND57	9220	14609	4780	4389	3080	31101	228
ND11	7005	71365	3620	9843	3840	64507	42
ND85	70248	99445	5209	11141	5067	52961	145
ND18	17725	17283	35802	21369	4242	4966	21
ND134	27465	23204	36161	4164	19322	22909	536

Prilog 7.7. Optimizacija hiperparametara modela mašinskog učenja.

Tabela P2. Lista hiperparametara sa vrednostima i ciljevima razmatranim pri optimizaciji modela mašinskog učenja.

Model	Algoritam	Optimizovani hiperparametri
Regresioni model	RF	{'bootstrap': [True, False], 'max_depth': [5, 20, 35, 50, 65, 80, None], 'max_features': ['log2', 'sqrt'], 'min_samples_leaf': [2, 4, 6], 'min_samples_split': [2, 5, 10], 'n_estimators': [250, 500, 750, 1000, 1250, 1500]}, scoring : (' r2')
	SVR	{'C': [0.01, 0.1, 1, 10, 100, 200, 400, 1000], 'gamma': [100, 10, 1, 1e-1, 1e-2, 1e-3], 'kernel': ['rbf', 'linear']}, scoring : (' r2')
	XGBoost	{'learning_rate': [0.01, 0.1], 'max_depth': [3, 6, 10], 'min_child_weight': [1, 3, 5], 'gamma': [0, 1, 5], 'subsample': [0.6, 0.7, 0.8, 0.9, 1.0], 'colsample_bytree': [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 'n_estimators': [500, 1500]}, scoring : (' r2')
	KNN	{'n_neighbors': [5, 9, 15, 19, 25, 35, 45, 55, 71], 'weights': ['distance'], 'p': [1, 2], 'metric': ['jaccard', 'dice', 'rogerstanimoto']}, scoring : (' r2')
	DNN	{number of Dense layers: (1-8, activation = 'relu'; None in output layer), number of neurons in each layer: ((100 - 5000), step = (50 - 1000)), added dropouts: ((0 - 0.5), step = 0.1), SGD optimizer learning rate: (1e-4 to 1e-2, sampling = ('log', default = 1e-3)),

		loss_fn : 'MeanSquaredError', batch_size : (32-128, step = 32, default = 64), objective : ('val_mean_squared_error','min'), used 30 % of training set as validation set }
Binarni Klasifikacioni model	RF	{'bootstrap': [True, False], 'max_depth': [5, 20, 35, 50, 65, 80, None], 'max_features': ['log2', 'sqrt'], 'min_samples_leaf': [2, 4, 6], 'min_samples_split': [2, 5, 10], 'n_estimators': [250,500,750,1000,1250,1500]}, scoring : ('balanced_accuracy')
	SVC	{'C': [0.01, 0.1, 1, 10, 100, 200, 400, 1000], 'gamma': [100,10,1,1e-1,1e-2, 1e-3], 'kernel': ['linear','rbf']}, scoring : ('balanced_accuracy')
	XGBoost	{'learning_rate': [0.01, 0.1], 'max_depth': [3, 6, 10], 'min_child_weight': [1, 3, 5], 'gamma': [0, 1, 5], 'subsample': [0.6, 0.7, 0.8, 0.9, 1.0], 'colsample_bytree': [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 'n_estimators': [500, 1500]}, scoring : ('balanced_accuracy')
	KNN	{'n_neighbors': [5,9,15,19,25,35,45,55,71], 'weights': ['distance'],'p': [1,2],'metric': ['jaccard', 'dice', 'rogerstanimoto']}, scoring : ('balanced_accuracy')
	DNN	{number of Dense layers: (1-8, (activation = 'relu'; 'sigmoid' in output layer)), number of neurons in each layer: ((100 - 5000), step = (50 - 1000)), added dropouts: ((0-0.5), step = 0.1), Adam optimizer learning rate: ((1e-4 to 1e-2), sampling = 'log', default = 1e-3), loss_fn : 'BinaryCrossentropy', batch_size : (32-128, step = 32, default = 64), objective : ('val_binary_accuracy','max'), used 30 % of training set as validation set }
Višeklasni Klasifikacioni model	RF	{'bootstrap': [True, False], 'max_depth': [5, 20, 35, 50, 65, 80, None], 'max_features': ['log2', 'sqrt'], 'min_samples_leaf': [2, 4, 6], 'min_samples_split': [2, 5, 10], 'n_estimators': [250,500,750,1000,1250,1500]}, scoring : ('balanced_accuracy')
	SVC	{'C': [0.01, 0.1, 1, 10, 100, 200, 400, 1000], 'gamma': [100,10,1,1e-1,1e-2, 1e-3], 'kernel': ['linear', 'rbf']}, scoring : ('balanced_accuracy')
	XGBoost	{'learning_rate': [0.01, 0.1], 'max_depth': [3, 6, 10], 'min_child_weight': [1, 3, 5], 'gamma': [0, 1, 5], 'subsample': [0.6, 0.7, 0.8, 0.9, 1.0], 'colsample_bytree': [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 'n_estimators': [500, 1500]}, scoring : ('balanced_accuracy')
	KNN	{'n_neighbors': [5,9,15,19,25,35,45,55,71], 'weights': ['distance'],'p': [1,2],'metric': ['jaccard', 'dice', 'rogerstanimoto']}, scoring : ('balanced_accuracy')
	DNN	{number of Dense layers: ((1-8), activation = ('relu'; 'softmax' in output layer)), number of neurons in each layer: ((100 - 5000), step = (50 - 1000)), added dropouts: (0 -0.5, step = 0.1), Adam optimizer learning rate: (1e-4 to 1e-2, sampling = ('log', default = 1e-3)), loss_fn : 'SparseCategoricalCrossentropy', batch_size : ((32-128), step = 32, default = 64), objective : ('val_categorical_accuracy','max'), used 30 % of training set as validation set }

Tabela P3. Vrednosti optimizovanih hiperparametara finalnih modela.

Finalni model	Optimizovani hiperparametri
Regresioni XGBoost:ECFP4	[('colsample_bytree', 0.5), ('gamma', 0), ('learning_rate', 0.01), ('max_depth', 10), ('min_child_weight', 3), ('n_estimators', 1500), ('subsample', 0.6)]
Binarni RF:ECFP4	[('bootstrap', False), ('max_depth', 65), ('max_features', 'sqrt'), ('min_samples_leaf', 2), ('min_samples_split', 5), ('n_estimators', 1000)]
Višeklasni SIRT1/2 RF:ECFP4	[('bootstrap', False), ('max_depth', 80), ('max_features', 'sqrt'), ('min_samples_leaf', 2), ('min_samples_split', 5), ('n_estimators', 750)]
Višeklasni SIRT2/3 DNN:deskriptorski model	<p>InputLayer : 'config': {'batch_input_shape': (None, 316), 'units': 4000, 'activation': 'relu'}</p> <p>Dense Layer 1 : 'config': {'name': 'dense_1', 'units': 100, 'activation': 'relu'}</p> <p>Dropout Layer : 'config': {'name': 'dropout', 'rate': 0.0}</p> <p>Dense Layer 2 : 'config': {'name': 'dense_2', 'units': 3, 'activation': 'softmax'}</p> <p>Optimizer: {'name': 'Adam', 'learning_rate': 1e-04, 'decay': 0.0, 'beta_1': 0.9, 'beta_2': 0.999, 'epsilon': 1e-07, 'amsgrad': False}</p> <p>batch_size = 32</p> <p>Loss function: SparseCategoricalCrossentropy</p>

Prilog 7.8. Rezultati interne i eksterne validacije modela mašinskog učenja.

Tabela P4. Rezultati interne (int), ukrštene (CV) i eksterne (ext) validacije regresionih modela.

Algoritam	Molekulska osobina	R^2_{int}	RMSE _{int}	Q^2	RMSE _{CV}	R^2_{ext}	RMSE _{ext}	r_m^2	\bar{r}_m^2	Δr_m^2	Q_{F1}^2	Q_{F2}^2	Q_{F3}^2	CCC	Kriterijum 1 ^a	Kriterijum 2 ^b	Kriterijum 3 ^c	k (ili k) ^d
RF	descriptors	0,99	0,11	0,7 +/- 0,02	0,55 +/- 0,01	0,7	0,55	0,7	0,52	0,27	0,7	0,7	0,7	0,81	0,008	0,300	0,206	1,000 (0,990)
	ECFP4	0,9	0,31	0,74 +/- 0,03	0,5 +/- 0,03	0,75	0,5	0,75	0,6	0,23	0,75	0,75	0,75	0,85	0,004	0,170	0,125	1,001 (0,991)
	MACCS	0,86	0,37	0,68 +/- 0,03	0,56 +/- 0,03	0,71	0,53	0,71	0,55	0,26	0,71	0,71	0,71	0,82	0,005	0,243	0,171	1,001 (0,991)
	ECFP6	0,93	0,26	0,75 +/- 0,03	0,49 +/- 0,03	0,77	0,48	0,77	0,62	0,21	0,77	0,77	0,76	0,86	0,003	0,139	0,104	1,001 (0,992)
SVM	descriptors	0,99	0,11	0,63 +/- 0,03	0,6 +/- 0,03	0,62	0,61	0,63	0,44	0,31	0,62	0,62	0,62	0,77	0,004	0,074	0,052	1,003 (0,989)
	ECFP4	0,98	0,15	0,74 +/- 0,03	0,51 +/- 0,03	0,74	0,51	0,74	0,63	0,13	0,74	0,74	0,73	0,84	0,004	0,476	0,296	0,997 (0,992)
	MACCS	0,96	0,21	0,66 +/- 0,05	0,57 +/- 0,04	0,68	0,57	0,68	0,55	0,21	0,68	0,68	0,68	0,81	0,001	0,173	0,116	1,002 (0,988)
	ECFP6	0,99	0,11	0,75 +/- 0,03	0,5 +/- 0,03	0,74	0,51	0,74	0,63	0,18	0,74	0,74	0,74	0,86	0,001	0,098	0,072	0,997 (0,995)
XGBoost	descriptors	0,99	0,09	0,7 +/- 0,02	0,55 +/- 0,01	0,67	0,58	0,66	0,53	0,25	0,68	0,68	0,68	0,82	0,001	0,089	0,065	0,999 (0,993)
	ECFP4*	0,97	0,17	0,74 +/- 0,02	0,5 +/- 0,02	0,79	0,46	0,79	0,7	0,17	0,74	0,74	0,74	0,86	9,06E-05	0,064	0,050	1,005 (0,998)
	MACCS	0,87	0,35	0,69 +/- 0,04	0,55 +/- 0,03	0,71	0,53	0,71	0,58	0,24	0,7	0,7	0,7	0,82	0,000	0,176	0,124	1,001 (0,990)
	ECFP6	0,99	0,1	0,73 +/- 0,02	0,52 +/- 0,02	0,73	0,52	0,73	0,62	0,2	0,73	0,73	0,73	0,87	0,000	0,115	0,084	0,993 (0,999)
KNN	descriptors	1	0,05	0,65 +/- 0,06	0,59 +/- 0,06	0,68	0,56	0,68	0,56	0,23	0,68	0,68	0,68	0,86	0,004	0,068	0,048	0,995 (0,997)
	ECFP4	1	0,07	0,7 +/- 0,04	0,54 +/- 0,04	0,74	0,51	0,75	0,64	0,13	0,74	0,74	0,74	0,87	0,000	0,180	0,123	0,986 (1,004)
	MACCS	0,97	0,17	0,6 +/- 0,07	0,63 +/- 0,05	0,6	0,63	0,6	0,47	0,16	0,6	0,6	0,6	0,79	0,014	0,216	0,123	0,991 (0,996)
	ECFP6	1	0,06	0,72 +/- 0,04	0,53 +/- 0,04	0,76	0,49	0,76	0,66	0,12	0,76	0,76	0,76	0,87	0,003	0,057	0,041	0,999 (0,993)

DNN	descriptors	0,98	0.16	0.79 +/- 0.07	0.45 +/- 0.06	0.66	0.58	0.68	0.57	0.03	0.66	0.66	0.66	0,81	0,032	0,053	0,015	1,005 (0,985)
	ECFP4	0,99	0.09	0.85 +/- 0.07	0.37 +/- 0.07	0.74	0.51	0.74	0.63	0.18	0.73	0.73	0.73	0,84	0,000	0,108	0,079	1,002 (0,989)
	MACCS	0,93	0.27	0.64+/-0.07	0.59+/-0.05	0.68	0.56	0.69	0.56	0.16	0.68	0.68	0.67	0,80	0,007	0,123	0,079	1,004 (0,985)
	ECFP6	0,96	0.19	0.82 +/- 0.05	0.41 +/- 0.07	0.73	0.52	0.74	0.63	0.17	0.73	0.73	0.73	0,81	0,048	0,571	0,380	1,170 (0,832)

* Finalna statistika modela

^a $(R^2 - R_0^2) / R^2 < 0.1$

^b $(R^2 - R_0'^2) / R^2 < 0.1$

^c $|R^2 - R_0'^2| < 0.3$

^d $0.85 \leq k \text{ (or } k') \leq 1.15$

Tabela P5. Rezultati interne (int) i ukrštene validacije (CV) binarnih klasifikacionih modela.

Algoritam	Molekulska osobina	BA _{int}	MCC _{int}	ROC_AUC _{int}	BA _{cv}	MCC _{cv}	ROC_AUC _{cv}
RF	ECFP4	0,95	0,89	0,99	0,80 +/- 0,04	0,60 +/- 0,04	0,90 +/- 0,02
	MACCS2	0,93	0,84	0,98	0,79 +/- 0,03	0,60 +/- 0,03	0,88 +/- 0,02
	descriptors	1	1	1	0,81 +/- 0,02	0,63 +/- 0,02	0,90 +/- 0,02
	ECFP6	0,97	0,93	1	0,80 +/- 0,04	0,62 +/- 0,04	0,90 +/- 0,02
SVC	ECFP4	1	0,99	1	0,78 +/- 0,04	0,59 +/- 0,04	0,87 +/- 0,03
	MACCS2	0,96	0,91	0,98	0,77 +/- 0,02	0,56 +/- 0,02	0,84 +/- 0,02
	descriptors	1	1	1	0,80 +/- 0,03	0,61 +/- 0,03	0,87 +/- 0,02
	ECFP6	1	1	1	0,67 +/- 0,03	0,44 +/- 0,03	0,87 +/- 0,02
XGBoost	ECFP4	0,99	0,99	1	0,77 +/- 0,03	0,55 +/- 0,03	0,88 +/- 0,02
	MACCS2	0,93	0,84	0,98	0,77 +/- 0,02	0,56 +/- 0,02	0,87 +/- 0,01
	descriptors	1	1	1	0,80 +/- 0,02	0,61 +/- 0,02	0,89 +/- 0,01
	ECFP6	1	1	1	0,79 +/- 0,03	0,58 +/- 0,03	0,88 +/- 0,02
KNN	ECFP4	1	0,99	1	0,79 +/- 0,03	0,59 +/- 0,03	0,87 +/- 0,03
	MACCS2	0,96	0,92	1	0,77 +/- 0,02	0,55 +/- 0,02	0,84 +/- 0,01
	descriptors	1	1	1	0,64 +/- 0,03	0,34 +/- 0,03	0,82 +/- 0,03
	ECFP6	1	1	1	0,80 +/- 0,03	0,61 +/- 0,03	0,88 +/- 0,03
DNN	ECFP4	1	0,99	1	0,8 +/- 0,05	0,61 +/- 0,05	0,89 +/- 0,02
	MACCS2	0,96	0,9	1	0,81 +/- 0,03	0,63 +/- 0,03	0,9 +/- 0,01
	descriptors	1	1	1	0,84 +/- 0,01	0,69 +/- 0,01	0,91 +/- 0,01
	ECFP6	0,98	0,96	1	0,82 +/- 0,02	0,64 +/- 0,02	0,91 +/- 0,02

Tabela P6. Rezultati interne (int) i ukrštene validacije (CV) višeklasnih klasifikacionih modela.

Model	Algoritam	Molecular feature	BA _{int}	MCC _{int}	ROC_AUC _{int}	BA _{CV}	MCC _{CV}	ROC_AUC _{CV}
SIRT1/2	RF	ECFP4	0,93	0,89	0,99	0,74 +/- 0,02	0,63 +/- 0,02	0,9 +/- 0,01
		MACCS2	0,9	0,85	0,99	0,73 +/- 0,03	0,62 +/- 0,03	0,89 +/- 0,01
		descriptors	1	1	1	0,75 +/- 0,03	0,63 +/- 0,03	0,9 +/- 0,01
		ECFP6	0,94	0,91	1	0,75 +/- 0,02	0,64 +/- 0,02	0,9 +/- 0,01
	SVC	ECFP4	0,97	0,95	1	0,75 +/- 0,03	0,64 +/- 0,03	0,89 +/- 0,01
		MACCS2	0,97	0,94	0,99	0,72 +/- 0,03	0,6 +/- 0,03	0,86 +/- 0,01
		descriptors	1	0,99	1	0,73 +/- 0,05	0,61 +/- 0,05	0,88 +/- 0,02
		ECFP6	1	1	1	0,75 +/- 0,03	0,64 +/- 0,03	0,89 +/- 0,01
	XGBoost	ECFP4	0,99	0,99	1	0,75 +/- 0,02	0,64 +/- 0,02	0,9 +/- 0,01
		MACCS2	0,92	0,88	0,99	0,74 +/- 0,01	0,63 +/- 0,01	0,9 +/- 0,01
		descriptors	1	1	1	0,75 +/- 0,02	0,64 +/- 0,02	0,89 +/- 0,01
		ECFP6	0,9	0,84	0,98	0,74 +/- 0,01	0,63 +/- 0,01	0,9 +/- 0,01
	KNN	ECFP4	1	1	1	0,72 +/- 0,01	0,59 +/- 0,01	0,88 +/- 0,01
		MACCS2	0,97	0,95	1	0,69 +/- 0,03	0,56 +/- 0,03	0,86 +/- 0,02
		descriptors	0,85	0,76	0,96	0,67 +/- 0,03	0,54 +/- 0,03	0,86 +/- 0,01
		ECFP6	1	1	1	0,71 +/- 0,02	0,58 +/- 0,02	0,89 +/- 0,01
DNN	ECFP4	1	1	1	0,75 +/- 0,03	0,63 +/- 0,02	0,9 +/- 0,02	
	MACCS2	0,95	0,92	1	0,74 +/- 0,01	0,61 +/- 0,05	0,89 +/- 0,01	
	descriptors	0,99	0,99	1	0,75 +/- 0,03	0,64 +/- 0,02	0,9 +/- 0,01	
	ECFP6	0,98	0,99	1	0,75 +/- 0,03	0,63 +/- 0,01	0,9 +/- 0,01	
SIRT2/3	RF	ECFP4	0,94	0,9	1	0,8 +/- 0,03	0,74 +/- 0,03	0,95 +/- 0,02
		MACCS2	0,94	0,89	0,99	0,78 +/- 0,01	0,72 +/- 0,01	0,93 +/- 0,02
		descriptors	1	1	1	0,8 +/- 0,03	0,74 +/- 0,03	0,94 +/- 0,02
		ECFP6	0,96	0,93	1	0,81 +/- 0,03	0,76 +/- 0,03	0,95 +/- 0,02
	SVC	ECFP4	1	1	1	0,71 +/- 0,03	0,65 +/- 0,03	0,94 +/- 0,03
		MACCS2	0,98	0,95	0,99	0,8 +/- 0,02	0,72 +/- 0,02	0,91 +/- 0,02
		descriptors	1	0,99	1	0,73 +/- 0,05	0,61 +/- 0,05	0,88 +/- 0,02
		ECFP6	1	1	1	0,78 +/- 0,03	0,7 +/- 0,03	0,94 +/- 0,02
	XGBoost	ECFP4	1	1	1	0,76 +/- 0,05	0,69 +/- 0,05	0,94 +/- 0,02
		MACCS2	0,98	0,95	1	0,78 +/- 0,04	0,7 +/- 0,04	0,92 +/- 0,02

		descriptors	1	1	1	0,79 +/- 0,01	0,72 +/- 0,01	0,94 +/- 0,01
		ECFP6	1	1	1	0,79 +/- 0,04	0,72 +/- 0,04	0,94 +/- 0,02
	KNN	ECFP4	1	1	1	0,77 +/- 0,05	0,7 +/- 0,05	0,92 +/- 0,03
		MACCS2	0,98	0,96	1	0,73 +/- 0,03	0,65 +/- 0,03	0,9 +/- 0,02
		descriptors	0,85	0,76	0,96	0,67 +/- 0,03	0,54 +/- 0,03	0,86 +/- 0,01
		ECFP6	1	1	1	0,77 +/- 0,04	0,69 +/- 0,04	0,92 +/- 0,03
	DNN	ECFP4	1	1	1	0,8 +/- 0,03	0,74 +/- 0,03	0,94 +/- 0,03
		MACCS2	1	1	1	0,78 +/- 0,03	0,74 +/- 0,03	0,94 +/- 0,02
		descriptors	1	1	1	0,8 +/- 0,03	0,76 +/- 0,03	0,95 +/- 0,02
		ECFP6	0,99	0,99	1	0,78 +/- 0,03	0,65 +/- 0,03	0,93 +/- 0,02

Tabela P7. Parametri eksterne validacije SIRT1/2 višeklasnih klasifikacionih modela.

Algoritam	Molekulska osobina	BA	MCC	ROC_AUC \bar{a}	Preciznost a	Opoziv ^a	F1 ^a
RF	descriptors	0,85	0,77	0,96	0,84	0,85	0,84
	ECFP4	0,84	0,74	0,96	0,81	0,84	0,82
	MACCS2	0,82	0,72	0,95	0,81	0,82	0,81
	ECFP6	0,83	0,74	0,96	0,81	0,83	0,82
SVC	descriptors	0,86	0,79	0,96	0,85	0,86	0,86
	ECFP4	0,83	0,73	0,95	0,81	0,83	0,82
	MACCS2	0,82	0,72	0,94	0,8	0,82	0,81
	ECFP6	0,8	0,7	0,95	0,8	0,8	0,8
XGBoost	descriptors	0,84	0,76	0,95	0,83	0,84	0,84
	ECFP4	0,84	0,74	0,96	0,82	0,84	0,83
	MACCS2	0,83	0,73	0,95	0,81	0,83	0,82
	ECFP6	0,81	0,71	0,95	0,79	0,82	0,8
KNN	descriptors	0,8	0,69	0,93	0,78	0,8	0,79
	ECFP4	0,84	0,73	0,94	0,82	0,84	0,82
	MACCS2	0,84	0,75	0,94	0,82	0,84	0,83
	ECFP6	0,83	0,71	0,93	0,81	0,83	0,81
DNN	descriptors	0,84	0,77	0,93	0,84	0,84	0,84
	ECFP4	0,82	0,77	0,94	0,84	0,82	0,83
	MACCS2	0,82	0,73	0,92	0,81	0,82	0,81
	ECFP6	0,84	0,74	0,95	0,82	0,84	0,82

^a Vrednosti su računane primenom makro-srednjih vrednosti i jedan-nasuprot-ostalnih (eng. *One vs rest*) pristupa.

Tabela P8. Statistički parametri SIRT1/2 višeklasnog klasifikacionog modela na validacionom setu sa molekulima mamcima.

Algoritam	Molekulske osobine	BA	MCC	ROC AU C ^a	Preciznost a	Opoziv ^a	F1 ^a
RF	descriptors	0,72	0.09	0.91	0.35	0.72	0.24
	ECFP4	0,82 (0,88)*	0.2 (0.73)*	0.92 (0.95)*	0.37 (0.72)*	0.82 (0.89)*	0.37 (0.79)*
	MACCS2	0,71	0.09	0.89	0.35	0.71	0.25
	ECFP6	0,8	0.16	0.91	0.36	0.8	0.34
SVC	descriptors	0,68	0.08	0.92	0.35	0.68	0.19
	ECFP4	0,7	0.09	0.9	0.35	0.7	0.23
	MACCS2	0,72	0.1	0.9	0.35	0.72	0.26
	ECFP6	0,69	0.08	0.91	0.35	0.69	0.22
XGBoost	descriptors	0,71	0.09	0.89	0.34	0.71	0.22
	ECFP4	0,75	0.12	0.93	0.35	0.75	0.28
	MACCS2	0,7	0.09	0.89	0.34	0.7	0.24
	ECFP6	0,73	0.11	0.91	0.35	0.73	0.28
KNN	descriptors	0,62	0.06	0.7	0.34	0.62	0.17
	ECFP4	0,69	0.08	0.81	0.34	0.69	0.22
	MACCS2	0,65	0.07	0.76	0.34	0.65	0.17
	ECFP6	0,67	0.08	0.81	0.34	0.67	0.21
DNN	descriptors	0,7	0.09	0.71	0.35	0.7	0.22
	ECFP4	0,67	0.08	0.87	0.34	0.67	0.22
	MACCS2	0,67	0.08	0.74	0.34	0.67	0.21
	ECFP6	0,68	0.08	0.87	0.34	0.68	0.21

^a Vrednosti su računane primenom makro-srednjih vrednost i jedan-nasuprot-ostalnih (eng. *One vs rest*) pristupa.* Vrednosti unutar zagrada su dobijene nakon korekcije domena primenjivosti.

Tabela P9. Parametri eksterne validacije SIRT2/3 višeklasnih klasifikacionih modela.

Algoritam	Molekulske osobine	BA	MCC	ROC $\frac{AUC}{a}$	Preciznost a	Opoziv a	F1 a
RF	descriptors	0,9	0,83	0,97	0,87	0,9	0,88
	ECFP4	0,83	0,73	0,96	0,8	0,83	0,81
	MACCS2	0,85	0,75	0,95	0,82	0,85	0,84
	ECFP6	0,83	0,74	0,95	0,8	0,83	0,81
SVC	descriptors	0,94	0,88	0,98	0,91	0,94	0,92
	ECFP4	0,77	0,71	0,97	0,83	0,78	0,8
	MACCS2	0,85	0,74	0,95	0,81	0,85	0,83
	ECFP6	0,82	0,74	0,95	0,82	0,82	0,82
XGBoost	descriptors	0,88	0,81	0,97	0,86	0,88	0,87
	ECFP4	0,84	0,73	0,96	0,81	0,85	0,82
	MACCS2	0,87	0,77	0,95	0,84	0,87	0,85
	ECFP6	0,8	0,71	0,96	0,79	0,81	0,8
KNN	descriptors	0,85	0,73	0,96	0,8	0,85	0,82
	ECFP4	0,85	0,74	0,93	0,8	0,85	0,82
	MACCS2	0,85	0,73	0,93	0,8	0,85	0,82
	ECFP6	0,85	0,73	0,93	0,79	0,85	0,82
DNN	descriptors	0,94	0,89	0,97	0,92	0,94	0,93
	ECFP4	0,84	0,75	0,95	0,81	0,84	0,83
	MACCS2	0,87	0,78	0,95	0,83	0,87	0,85
	ECFP6	0,88	0,81	0,97	0,85	0,88	0,86

^a Vrednosti su računane primenom makro-srednjih vrednosti i jedan-nasuprot-ostalnih (eng. *One vs rest*) pristupa.

Tabela P10. Statistički parametri SIRT2/3 višeklasnog klasifikacionog modela na validacionom setu sa molekulima mamcima.

Algoritam	Molekulske osobine	BA	MCC	ROC_AUC ^a	Preciznost ^a	Opoziv ^a	F1 ^a
RF	descriptors	0,66 (0,59)*	0.06 (0.12)*	0.93 (0.88)*	0.34 (0.47)*	0.66 (0.59)*	0.14 (0.34)*
	ECFP4	0,59	0.08	0.91	0.34	0.59	0.27
	MACCS2	0,65	0.06	0.92	0.34	0.65	0.17
	ECFP6	0,62	0.08	0.92	0.34	0.62	0.25
SVC	descriptors	0,62	0.06	0.95	0.34	0.62	0.02
	ECFP4	0,54	0.04	0.92	0.32	0.54	0.06
	MACCS2	0,58	0.05	0.93	0.34	0.58	0.08
	ECFP6	0,62	0.05	0.92	0.34	0.62	0.11
XGBoost	descriptors	0,7	0.07	0.92	0.34	0.7	0.19
	ECFP4	0,61	0.07	0.91	0.35	0.61	0.17
	MACCS2	0,69	0.07	0.92	0.34	0.69	0.21
	ECFP6	0,63	0.07	0.92	0.35	0.63	0.17
KNN	descriptors	0,64	0.05	0.71	0.34	0.64	0.15
	ECFP4	0,66	0.06	0.74	0.34	0.66	0.17
	MACCS2	0,61	0.04	0.66	0.34	0.61	0.12
	ECFP6	0,65	0.06	0.73	0.34	0.65	0.16
DNN	descriptors	0,72 (0,72)*	0.07 (0.1)*	0.87 (0.87)*	0.34 (0.34)*	0.72 (0.75)*	0.18 (0.2)*
	ECFP4	0,65	0.06	0.87	0.34	0.65	0.16
	MACCS2	0,67	0.06	0.87	0.34	0.67	0.18
	ECFP6	0,63	0.06	0.8	0.34	0.63	0.1

^a Vrednosti su računane primenom makro-srednjih vrednosti i jedan-nasuprot-ostalnih (eng. *One vs rest*) pristupa. * Vrednosti unutar zagrada su dobijene nakon korekcije domena primenjivosti.

Prilog 7.9. Rezultati eksperimentalne validacije *SIRT2i_Predictor-a*.

Tabela P11. Primenjivost *SIRT2i_Predictor-a* u analizi rezultata prospektivnog skrininga izvedenog primenom multistrukturalnog SBVS modela.

Jedinjenje	SMILES	Aktivnost (Inh% ili IC50 (µM)) ^a	SIRT2 aktivnost na osnovu kriterijuma iz poglavlja 3.1.5	Predviđanja binarnog modela ^b	Predviđene verovatnoće (binarni model) ^c
ND134*	<chem>CCN1C(=O)c2cccc3c(S(=O)(=O)NCCc4cccc4)ccc1c23</chem>	70,2%*	Siva zona	Izvan AD	0,455224
ND18	<chem>O=C(CCc1cccc1)Nc1ccc(Cc2cnc2)cc1</chem>	58,7 µM	Siva zona	Ne	0,354493
ND85	<chem>Cc1cccc1C(=O)NC(=S)NCc1ccc(-c2nn3c(C)nnc3s2)cc1</chem>	85,9 µM	Siva zona	Izvan AD	0,555415
ND11	<chem>Cc1cccc1OCC(=O)Nc1ccc(Oc2ccc3c(c2)C(=O)NC3=O)cc1</chem>	55,6%	Siva zona	Ne	0,383697
ND57	<chem>Cc1nnc2sc(-c3ccc(CNC(=O)c4ccc(-c5ccc(Cl)cc5)o4)cc3)nn12</chem>	47,2%	Siva zona	Izvan AD	0,415945
ND20	<chem>O=C(NCCc1cccc1)c1ccc(NC(=O)c2cc(-n3cnc3)ccc2Cl)cc1</chem>	45,1%	Siva zona	Ne	0,335031
ND10	<chem>Cc1ccc(C)c(SCCc2nn3c(Cc4cccc4F)nnc3s2)c1</chem>	40,5%	Siva zona	Izvan AD	0,549759
ND4	<chem>Cc1cc(=O)oc2cc(OC[C@H](O)Cn3c(C)nc4cccc43)cc1c2</chem>	31,8%	Neaktivni	Izvan AD	0,518789
ND26	<chem>O=C1Nc2ccc(S(=O)(=O)N3CCN(c4cccc4)CC3)c3ccc1c23</chem>	30,0%	Neaktivni	Ne	0,391046

^a Inh% testiranja su izvedena pri koncentraciji jedinjenja od 200 µM.

^b Ne – nisu predviđena kao aktivni SIRT2 inhibitori.

^c Izvan domena primenjivosti (AD) – Ukoliko je verovatnoća predviđanja između 0,4 i 0,6.

* Jedinjenje interferira sa fluorescentnim esejom.

Tabela P12. Rezultati *in vitro* evaluacije baze dostupnih jedinjenja.

Jedinjenje	SMILES	Aktivnost (Inh%) ^a	SIRT2 aktivnost na osnovu kriterijuma iz poglavlja 3.1.5	Predviđanja binarnog modela ^b	Predviđene verovatnoće (binarni model) ^c
316	<chem>COc1ccc([C@H]2Cc3cc(OC4O[C@H](CO)[C@@H](O)[C@H](O)[C@H]4O)cc(O)c3C(=O)O2)cc1</chem>	28,2%	Neaktivni	Izvan AD	0,473878
1024*	†	48,7% *	Siva zona	Ne	0,294624
1558	<chem>COc1ccc(Cn2cnc3c(Cl)nccc32)cc1</chem>	51,1%	Siva zona	Ne	0,34941
1657	<chem>COc1ccc(CN2NC([N+](=O)[O-])c3ccnc(NC4CCCC4)c32)cc1</chem>	59,7%	Siva zona	Izvan AD	0,427429
EX-527	<chem>NC(=O)C1CCCc2c1[nH]c1ccc(Cl)cc21</chem>	89,6%	Aktivni	Da	0,830218

^a Inh% testiranja su izvedena pri koncentraciji jedinjenja od 200 µM
^b Ne – nisu predviđena kao aktivni SIRT2 inhibitori; Da – predviđeni kao aktivni SIRT2 inhibitori.
^c Izvan domena primenjivosti (AD) – Ukoliko je verovatnoća predviđanja između 0,4 i 0,6.
* Jedinjenje testirano pri koncentraciji od 100 µM usled interferencija.
† Struktura jedinjenja dostupna na zahtev.

Prilog 7.10. NMR (^1H i ^{13}C) spektri prekursora i finalnih jedinjenja.

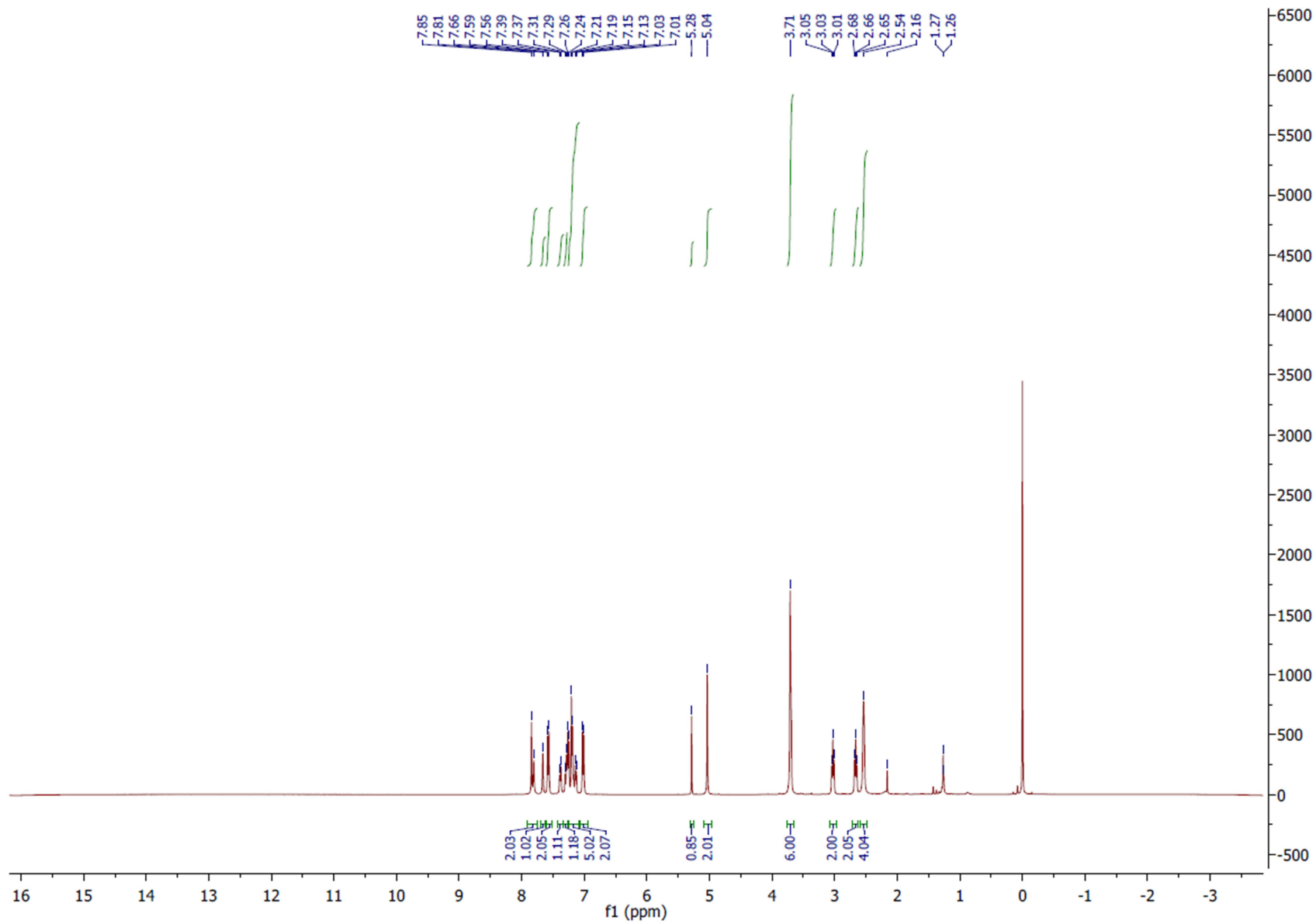
Tabela P13. Sažetak ^1H i ^{13}C spektara prekursora i finalnih jedinjenja.

ND3_2	<p>^1H NMR (400 MHz, CDCl_3) δ 8.31 (s, 1H), 8.18 (d, $J = 8.1$ Hz, 1H), 7.76 (d, $J = 7.7$ Hz, 1H), 7.56 (t, $J = 7.9$ Hz, 1H), 7.34 – 7.26 (m, 2H), 6.91 (d, $J = 8.9$ Hz, 2H), 6.37 (s, 1H), 5.12 (s, 2H), 1.51 (s, 9H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 154.14 (s), 153.13 (s), 148.45 (s), 139.35 (s), 133.11 (s), 132.30 (s), 129.55 (s), 122.89 (s), 122.15 (s), 120.55 (s), 115.33 (s), 80.43 (s), 69.05 (s), 28.36 (s).</p>
ND3_4	<p>^1H NMR (400 MHz, CDCl_3) δ 7.44 (t, $J = 12.5$ Hz, 3H), 7.27 (d, $J = 3.1$ Hz, 1H), 7.22 (dd, $J = 15.8, 8.3$ Hz, 6H), 7.10 (d, $J = 7.4$ Hz, 1H), 6.89 – 6.78 (m, 2H), 6.43 (s, 1H), 4.93 (d, $J = 9.2$ Hz, 2H), 3.01 (t, $J = 7.6$ Hz, 2H), 2.61 (t, $J = 7.7$ Hz, 2H), 1.50 (s, 9H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 170.63 (s), 154.75 (s), 153.30 (s), 140.64 (s), 138.11 (s), 138.06 (s), 131.63 (s), 129.16 (s), 128.61 (s), 128.37 (s), 126.34 (s), 123.00 (s), 120.80 (s), 119.42 (s), 118.66 (s), 115.26 (s), 80.37 (s), 69.95 (s), 39.30 (s), 31.51 (s), 28.37 (s).</p>
ND3_5	<p>^1H NMR (400 MHz, CDCl_3) δ 7.50 (s, 1H), 7.45 – 7.37 (m, 2H), 7.24 (t, $J = 7.6$ Hz, 2H), 7.18 (t, $J = 6.6$ Hz, 3H), 7.10 (d, $J = 7.5$ Hz, 1H), 6.75 (d, $J = 8.7$ Hz, 2H), 6.58 (d, $J = 8.7$ Hz, 2H), 4.88 (s, 2H), 3.18 (d, $J = 63.5$ Hz, 2H), 2.99 (t, $J = 7.6$ Hz, 2H), 2.59 (t, $J = 7.7$ Hz, 2H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 170.63 (s), 151.81 (s), 140.63 (s), 140.24 (s), 138.46 (s), 138.02 (s), 129.10 (s), 128.60 (s), 128.35 (s), 126.33 (s), 123.10 (s), 119.37 (s), 118.78 (s), 116.39 (s), 116.01 (s), 70.37 (s), 39.25 (s), 31.48 (s).</p>
ND3_6	<p>^1H NMR (400 MHz, CDCl_3) δ 7.55 (s, 1H), 7.40 (d, $J = 12.6$ Hz, 1H), 7.36 (d, $J = 8.0$ Hz, 1H), 7.27 (t, $J = 7.5$ Hz, 3H), 7.19 (d, $J = 7.2$ Hz, 3H), 7.12 (d, $J = 7.4$ Hz, 1H), 6.94 – 6.86 (m, 4H), 4.95 (s, 2H), 3.01 (t, $J = 7.6$ Hz, 2H), 2.63 (t, $J = 7.6$ Hz, 2H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 170.65 (s), 155.98 (s), 140.55 (s), 138.11 (s), 137.78 (s), 132.67 (s), 129.23 (s), 128.64 (s), 128.36 (s), 126.40 (s), 123.10 (s), 120.02 (s), 119.49 (s), 118.77 (s), 116.16 (s), 70.05 (s), 39.34 (s), 31.50 (s).</p>
ND3_7	<p>^1H NMR (400 MHz, CDCl_3) δ 7.83 (d, $J = 14.4$ Hz, 2H), 7.66 (s, 1H), 7.58 (d, $J = 8.7$ Hz, 2H), 7.38 (d, $J = 7.8$ Hz, 1H), 7.30 (d, $J = 7.9$ Hz, 1H), 7.26 – 7.08 (m, 5H), 7.02 (d, $J = 8.8$ Hz, 2H), 5.28 (s, 1H), 5.04 (s, 2H), 3.71 (s, 6H), 3.03 (t, $J = 7.6$ Hz, 2H), 2.66 (t, $J = 7.6$ Hz, 2H), 2.54 (s, 4H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 170.42 (s), 158.79 (s), 144.73 (s), 140.56 (s), 138.17 (s), 137.40 (s), 130.80 (s), 129.34 (s), 128.69 (s), 128.40 (s), 126.45 (s), 123.08 (s), 122.14 (s), 120.98 (s), 119.47 (s), 118.70 (s), 115.78 (s), 70.09 (s), 66.89 (s), 53.72 (s), 53.49 (s), 39.49 (s), 31.51 (s).</p>
ND4_1	<p>^1H NMR (400 MHz, CDCl_3) δ 7.35 (s, 1H), 7.05 (dd, $J = 16.1, 8.0$ Hz, 2H), 6.85 – 6.74 (m, 2H), 6.54 (dd, $J = 8.1, 1.5$ Hz, 1H), 1.48 (s, 9H).</p>
ND4_2	<p>^1H NMR (400 MHz, CDCl_3) δ 8.30 (s, 1H), 8.16 (d, $J = 8.2$ Hz, 1H), 7.75 (d, $J = 7.6$ Hz, 1H), 7.54 (t, $J = 7.9$ Hz, 1H), 7.31 (s, 1H), 7.18 (t, $J = 8.1$ Hz, 1H), 6.82 (d, $J = 8.0$ Hz, 1H), 6.68 – 6.58 (m, 2H), 5.12 (s, 2H), 1.52 (s, 9H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 158.84 (s), 152.63 (s), 148.37 (s), 139.85 (s), 139.20 (s), 133.20 (s), 129.81 (s), 129.51 (s), 122.84 (s), 122.16 (s), 111.37 (s), 109.49 (s), 104.94 (s), 80.66 (s), 68.54 (s), 28.32 (s).</p>
ND4_3	<p>^1H NMR (400 MHz, CDCl_3) δ 7.26 – 7.08 (m, 3H), 6.77 (dt, $J = 34.5, 13.3$ Hz, 3H), 6.60 (d, $J = 8.7$ Hz, 3H), 4.93 (s, 2H), 3.68 (s, 2H), 1.51 (s, 9H).</p> <p>^{13}C NMR (101 MHz, CDCl_3) δ 159.46 (s), 152.67 (s), 146.66 (s), 139.61 (s), 138.16 (s), 129.64 (s), 129.43 (s), 117.49 (s), 114.61 (s), 114.01 (s), 110.93 (s), 109.61 (s), 105.09 (s), 80.48 (s), 69.89 (s), 28.33 (s).</p>
ND4_4	<p>^1H NMR (400 MHz, CDCl_3) δ 7.61 – 7.49 (m, 2H), 7.38 (s, 1H), 7.31 – 7.04 (m, 9H), 6.82 (d, $J = 8.0$ Hz, 1H), 6.66 – 6.53 (m, 2H), 4.94 (s, 2H), 3.04 (t, $J = 7.6$ Hz, 2H), 2.67 (t, $J = 7.7$ Hz, 2H), 1.51 (s, 9H).</p>

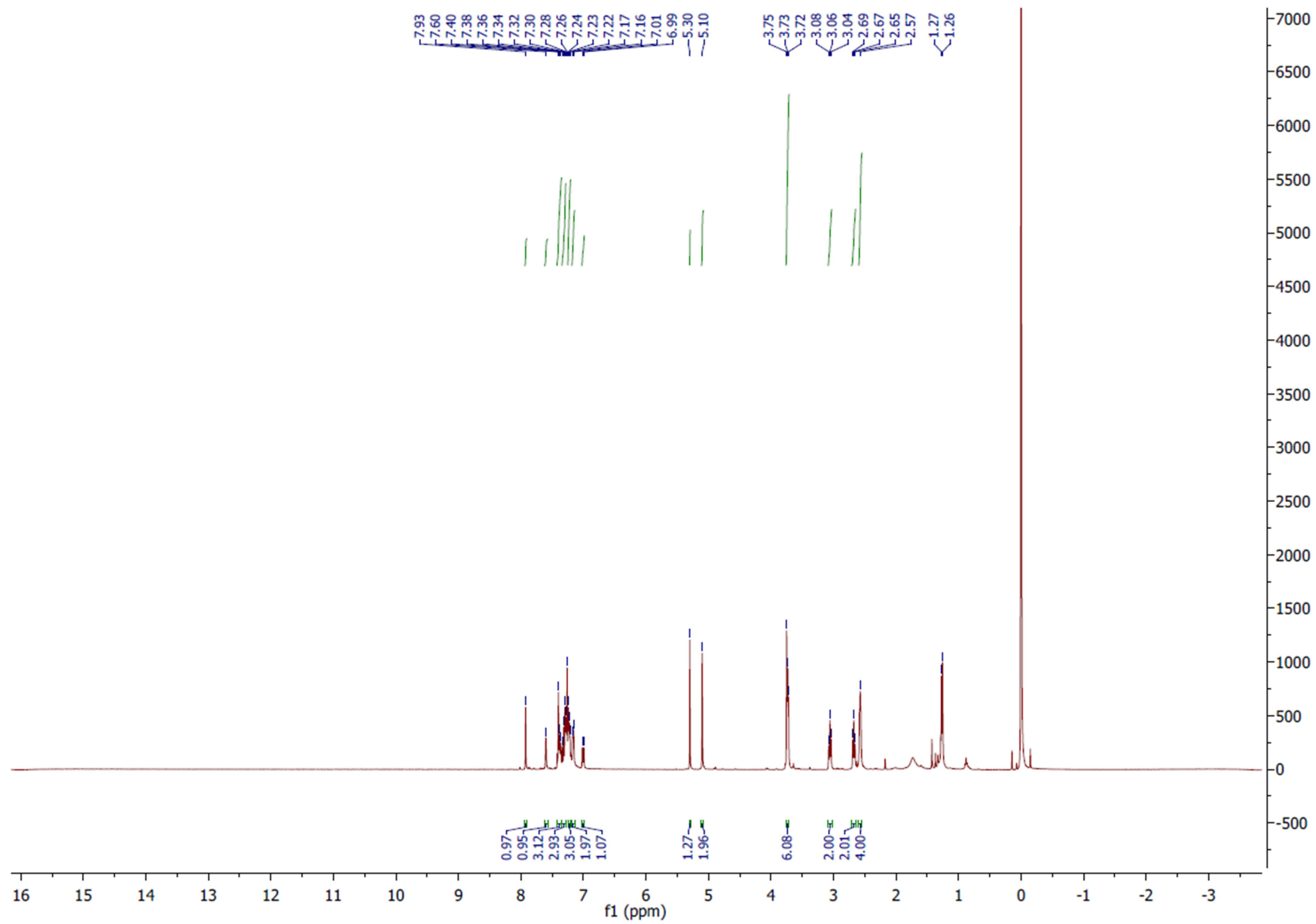
	¹³ C NMR (101 MHz, CDCl ₃) δ 170.68 (s), 159.26 (s), 152.76 (s), 140.66 (s), 139.55 (s), 138.12 (s), 137.86 (s), 129.68 (s), 129.13 (s), 128.60 (s), 128.38 (s), 126.33 (s), 122.92 (s), 119.38 (s), 118.56 (s), 111.13 (s), 109.32 (s), 105.45 (s), 80.58 (s), 69.43 (s), 39.35 (s), 31.53 (s), 28.34 (s).
ND4_5	¹ H NMR (400 MHz, CDCl ₃) δ 7.51 (s, 1H), 7.46 – 7.37 (m, 2H), 7.31 – 7.13 (m, 6H), 7.09 (d, <i>J</i> = 7.5 Hz, 1H), 7.01 (t, <i>J</i> = 8.0 Hz, 1H), 6.33 (dd, <i>J</i> = 8.2, 1.4 Hz, 1H), 6.25 (dd, <i>J</i> = 8.1, 4.9 Hz, 2H), 4.90 (s, 2H), 3.46 (s, 2H), 2.99 (t, <i>J</i> = 7.6 Hz, 2H), 2.59 (t, <i>J</i> = 7.7 Hz, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 170.69 (s), 159.80 (s), 147.83 (s), 140.63 (s), 138.19 (s), 138.06 (s), 130.12 (s), 129.13 (s), 128.60 (s), 128.36 (s), 126.33 (s), 123.03 (s), 119.43 (s), 118.75 (s), 108.23 (s), 104.84 (s), 101.95 (s), 69.41 (s), 39.22 (s), 31.47 (s).
ND4_6	¹ H NMR (400 MHz, CDCl ₃) δ 7.55 (d, <i>J</i> = 15.3 Hz, 1H), 7.42 – 7.06 (m, 11H), 6.73 – 6.58 (m, 2H), 5.20 – 4.80 (m, 2H), 3.09 – 2.95 (m, 2H), 2.64 (t, <i>J</i> = 7.5 Hz, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 170.49 (s), 159.82 (s), 141.31 (s), 140.57 (s), 138.11 (s), 137.58 (s), 130.50 (s), 129.28 (s), 128.66 (s), 128.38 (s), 126.42 (s), 123.12 (s), 119.50 (s), 118.72 (s), 111.65 (s), 111.38 (s), 106.01 (s), 69.79 (s), 39.42 (s), 31.51 (s).
ND4_7	¹ H NMR (400 MHz, CDCl ₃) δ 7.93 (s, 1H), 7.60 (s, 1H), 7.43 – 7.35 (m, 3H), 7.35 – 7.28 (m, 3H), 7.25 – 7.21 (m, 3H), 7.17 (d, <i>J</i> = 5.4 Hz, 2H), 7.00 (d, <i>J</i> = 8.5 Hz, 1H), 5.30 (s, 1H), 5.10 (s, 2H), 3.76 – 3.71 (m, 6H), 3.06 (t, <i>J</i> = 7.6 Hz, 2H), 2.67 (t, <i>J</i> = 7.6 Hz, 2H), 2.57 (s, 4H). ¹³ C NMR (101 MHz, CDCl ₃) δ 170.83 (s), 159.59 (s), 145.04 (s), 140.56 (s), 138.18 (s), 138.08 (s), 137.29 (s), 130.62 (s), 129.35 (s), 128.68 (s), 128.40 (s), 126.44 (s), 123.14 (s), 120.92 (s), 119.54 (s), 118.74 (s), 115.21 (s), 112.68 (s), 107.38 (s), 70.04 (s), 66.88 (s), 53.68 (s), 53.48 (s), 39.49 (s), 31.51 (s, <i>J</i> = 27.0 Hz).
ND5_4	¹ H NMR (400 MHz, CDCl ₃) δ 7.89 (s, 1H), 7.73 (d, <i>J</i> = 15.5 Hz, 1H), 7.62 (s, 2H), 7.48 (d, <i>J</i> = 2.7 Hz, 2H), 7.38 – 7.28 (m, 4H), 7.22 (d, <i>J</i> = 8.2 Hz, 2H), 7.12 (t, <i>J</i> = 12.0 Hz, 1H), 6.83 (t, <i>J</i> = 10.5 Hz, 2H), 6.59 (t, <i>J</i> = 11.6 Hz, 1H), 6.40 (s, 1H), 4.95 (d, <i>J</i> = 19.8 Hz, 2H), 1.51 (s, 9H). ¹³ C NMR (101 MHz, CDCl ₃) δ 164.23 (s), 154.79 (s), 153.31 (s), 142.32 (s), 138.41 (s), 138.17 (s), 134.64 (s), 131.59 (s), 129.94 (s), 129.29 (s), 128.86 (s), 127.97 (s), 123.10 (s), 120.93 (s), 120.81 (s), 119.47 (s), 118.66 (s), 115.28 (s), 80.34 (s), 69.95 (s), 28.37 (s).
ND5_5	¹ H NMR (400 MHz, CDCl ₃) δ 7.75 (t, <i>J</i> = 13.5 Hz, 2H), 7.62 (d, <i>J</i> = 10.7 Hz, 2H), 7.55 – 7.19 (m, 8H), 7.15 (d, <i>J</i> = 7.5 Hz, 1H), 6.78 (d, <i>J</i> = 8.7 Hz, 2H), 6.61 (d, <i>J</i> = 8.7 Hz, 2H), 6.56 (d, <i>J</i> = 15.5 Hz, 1H), 4.94 (s, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 164.26 (s), 151.86 (s), 142.40 (s), 140.27 (s), 138.65 (s), 138.30 (s), 134.62 (s), 129.96 (s), 129.26 (s), 128.87 (s), 127.98 (s), 123.26 (s), 120.84 (s), 120.77 (s), 118.77 (s), 116.41 (s), 116.07 (s), 70.42 (s).
ND5_6	¹ H NMR (400 MHz, CDCl ₃) δ 7.79 – 7.71 (m, 2H), 7.68 (s, 1H), 7.56 (d, <i>J</i> = 7.6 Hz, 1H), 7.52 – 7.45 (m, 2H), 7.34 (dd, <i>J</i> = 13.8, 5.3 Hz, 4H), 7.16 (d, <i>J</i> = 7.5 Hz, 1H), 6.92 (s, 4H), 6.57 (d, <i>J</i> = 15.5 Hz, 1H), 5.00 (s, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 164.20 (s), 156.01 (s), 142.63 (s), 138.39 (s), 137.94 (s), 134.55 (s), 132.70 (s), 130.06 (s), 129.37 (s), 128.90 (s), 127.99 (s), 123.24 (s), 120.68 (s), 120.04 (s), 119.57 (s), 118.79 (s), 116.19 (s), 70.08 (s).
ND5_7	¹ H NMR (400 MHz, CDCl ₃) δ 7.86 (s, 1H), 7.83 (s, 1H), 7.78 (s, 1H), 7.74 (s, 1H), 7.61 (d, <i>J</i> = 8.8 Hz, 2H), 7.53 (d, <i>J</i> = 3.8 Hz, 3H), 7.37 (t, <i>J</i> = 7.1 Hz, 4H), 7.20 (d, <i>J</i> = 7.3 Hz, 1H), 7.07 (d, <i>J</i> = 8.8 Hz, 2H), 6.59 (d, <i>J</i> = 15.5 Hz, 1H), 5.12 (s, 2H), 3.77 – 3.68 (m, 6H), 2.57 (s, 4H). ¹³ C NMR (101 MHz, CDCl ₃) δ 186.01 (s), 158.82 (s), 151.15 (s), 145.08 (s), 142.63 (s), 138.45 (s), 137.49 (s), 134.58 (s), 130.75 (s), 130.08 (s), 129.43 (s), 128.92 (s), 127.99 (s), 123.37 (s), 122.16 (s), 121.08 (s), 120.69 (s), 118.62 (s), 115.80 (s), 70.11 (s), 66.86 (s), 53.70 (s), 53.47 (s).
ND6_4	¹ H NMR (400 MHz, CDCl ₃) δ 7.44 (d, <i>J</i> = 8.0 Hz, 1H), 7.43 – 7.36 (m, 3H), 7.36 – 7.27 (m, 4H), 7.23 (d, <i>J</i> = 9.0 Hz, 3H), 7.13 (d, <i>J</i> = 7.5 Hz, 1H), 6.86 (d, <i>J</i> = 8.9 Hz, 2H), 6.36 (s, 1H), 4.96 (s, 2H), 3.72 (s, 2H), 1.50 (s, 9H).

	¹³ C NMR (101 MHz, CDCl ₃) δ 169.14 (s), 154.68 (s), 153.17 (s), 138.09 (s), 137.93 (s), 134.36 (s), 131.71 (s), 129.53 (s), 129.25 (s), 129.20 (s), 127.70 (s), 123.30 (s), 120.59 (s), 119.31 (s), 118.56 (s), 115.27 (s), 80.28 (s), 69.99 (s), 44.84 (s), 28.37 (s).
ND6_5	¹ H NMR (400 MHz, CDCl ₃) δ 7.46 – 7.29 (m, 8H), 7.27 – 7.20 (m, 1H), 7.12 (d, <i>J</i> = 7.5 Hz, 1H), 6.67 (dd, <i>J</i> = 62.9, 8.7 Hz, 4H), 4.90 (s, 2H), 3.70 (s, 2H), 3.44 (s, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 169.19 (s), 151.81 (s), 140.30 (s), 138.53 (s), 137.90 (s), 134.41 (s), 129.50 (s), 129.20 (s), 129.12 (s), 127.64 (s), 123.35 (s), 119.28 (s), 118.67 (s), 116.36 (s), 116.05 (s), 70.43 (s), 44.78 (s).
ND6_6	¹ H NMR (400 MHz, CDCl ₃) δ 7.54 (s, 1H), 7.42 – 7.22 (m, 8H), 7.13 (d, <i>J</i> = 7.5 Hz, 1H), 6.95 – 6.86 (m, 4H), 4.96 (s, 2H), 3.72 (s, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 169.18 (s), 155.98 (s), 138.00 (s), 137.79 (s), 134.33 (s), 132.70 (s), 129.52 (s), 129.26 (s), 129.24 (s), 127.72 (s), 123.33 (s), 120.02 (s), 119.41 (s), 118.67 (s), 116.17 (s), 70.09 (s), 44.85 (s).
ND6_7	¹ H NMR (400 MHz, CDCl ₃) δ 7.84 (s, 1H), 7.65 – 7.57 (m, 3H), 7.43 – 7.37 (m, 2H), 7.36 – 7.26 (m, 6H), 7.16 (d, <i>J</i> = 6.9 Hz, 1H), 7.04 (d, <i>J</i> = 8.9 Hz, 2H), 5.06 (s, 2H), 3.73 (dd, <i>J</i> = 8.0, 4.1 Hz, 8H), 2.56 (s, 4H). ¹³ C NMR (101 MHz, CDCl ₃) δ 169.19 (s), 158.75 (s), 144.77 (s), 138.09 (s), 137.36 (s), 134.32 (s), 130.78 (s), 129.52 (s), 129.30 (s), 129.27 (s), 127.74 (s), 123.31 (s), 122.11 (s), 120.96 (s), 119.43 (d, <i>J</i> = 11.0 Hz), 118.65 (d, <i>J</i> = 10.6 Hz), 115.76 (s), 70.08 (s), 66.90 (s), 53.73 (s), 53.49 (s), 44.86 (s).
ND7_4	¹ H NMR (400 MHz, CDCl ₃) δ 7.76 (d, <i>J</i> = 15.6 Hz, 2H), 7.55 (d, <i>J</i> = 22.3 Hz, 3H), 7.41 – 7.29 (m, 4H), 7.26 (s, 1H), 7.17 (dd, <i>J</i> = 17.3, 9.6 Hz, 3H), 6.83 (d, <i>J</i> = 7.9 Hz, 1H), 6.60 (t, <i>J</i> = 12.3 Hz, 3H), 4.99 (d, <i>J</i> = 18.0 Hz, 2H), 1.52 (s, 9H).
ND7_5	¹ H NMR (400 MHz, CDCl ₃) δ 7.83 (s, 1H), 7.73 (d, <i>J</i> = 15.5 Hz, 1H), 7.62 (d, <i>J</i> = 10.9 Hz, 2H), 7.50 – 7.43 (m, 2H), 7.37 – 7.27 (m, 4H), 7.17 – 6.99 (m, 2H), 6.56 (d, <i>J</i> = 15.5 Hz, 1H), 6.38 – 6.32 (m, 1H), 6.31 – 6.22 (m, 2H), 4.96 (s, 2H), 3.57 (d, <i>J</i> = 71.3 Hz, 2H).
ND7_7	¹ H NMR (400 MHz, CDCl ₃) δ 7.94 (s, 1H), 7.76 (d, <i>J</i> = 15.4 Hz, 3H), 7.59 (s, 1H), 7.52 (d, <i>J</i> = 3.6 Hz, 2H), 7.42 – 7.28 (m, 7H), 7.19 (d, <i>J</i> = 7.6 Hz, 1H), 7.00 (d, <i>J</i> = 8.6 Hz, 1H), 6.61 (d, <i>J</i> = 15.5 Hz, 1H), 5.11 (s, 2H), 3.74 (dd, <i>J</i> = 9.8, 5.3 Hz, 6H), 2.58 (s, 4H). ¹³ C NMR (101 MHz, CDCl ₃) δ 164.16 (s), 159.57 (s), 144.85 (s), 142.57 (s), 138.52 (s), 138.02 (s), 137.37 (s), 134.57 (s), 130.62 (s), 130.04 (s), 129.43 (s), 128.89 (s), 127.98 (s), 123.23 (s), 121.02 (s), 120.78 (s), 119.67 (s), 118.85 (s), 115.24 (s), 112.67 (s), 107.41 (s), 70.03 (s), 66.82 (s), 53.64 (s), 53.44 (s).
ND18	¹ H NMR (400 MHz, CDCl ₃) δ 8.47 (d, <i>J</i> = 4.9 Hz, 2H), 7.44 – 7.00 (m, 12H), 3.91 (s, <i>J</i> = 23.7 Hz, 2H), 3.04 (t, <i>J</i> = 7.4 Hz, 2H), 2.65 (t, <i>J</i> = 7.5 Hz, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 170.39 (s), 150.05 (s), 149.81 (s), 140.63 (s), 136.48 (s), 134.73 (s), 129.54 (s), 128.65 (s), 128.40 (s), 126.40 (s), 124.15 (s), 120.29 (s), 40.63 (s), 39.41 (s), 31.57 (s).
ND19	¹ H NMR (400 MHz, CDCl ₃) δ 8.50 (d, <i>J</i> = 4.7 Hz, 2H), 7.82 – 7.04 (m, 14H), 6.57 (d, <i>J</i> = 15.5 Hz, 1H), 3.94 (s, 2H). ¹³ C NMR (101 MHz, CDCl ₃) δ 164.02 (s), 150.05 (s), 149.84 (s), 142.45 (s), 136.79 (s), 134.62 (s), 130.00 (s), 129.67 (s), 128.89 (s), 127.96 (s), 124.19 (s), 120.76 (s), 120.31 (s), 115.40 (s), 40.67 (s).
ND20	¹ H NMR (400 MHz, CDCl ₃) δ 8.44 (d, <i>J</i> = 4.0 Hz, 2H), 8.29 (s, 1H), 7.46 (dd, <i>J</i> = 20.4, 8.1 Hz, 4H), 7.28 (d, <i>J</i> = 8.2 Hz, 2H), 7.12 – 7.04 (m, 4H), 3.90 (s, 2H), 3.07 (t, <i>J</i> = 7.5 Hz, 2H), 2.65 (t, <i>J</i> = 7.5 Hz, 2H).
ND21	¹ H NMR (400 MHz, DMSO) δ 10.11 (s, 1H), 8.43 (d, <i>J</i> = 5.7 Hz, 2H), 7.52 (d, <i>J</i> = 8.4 Hz, 2H), 7.31 (d, <i>J</i> = 4.4 Hz, 3H), 7.18 (dd, <i>J</i> = 14.4, 7.0 Hz, 4H), 3.89 (s, 2H), 3.60 (s, 2H), 3.30 (s, <i>J</i> = 9.5 Hz, 2H). ¹³ C NMR (101 MHz, DMSO) δ 169.41 (s), 150.71 (s), 150.04 (s), 138.06 (s), 136.49 (s), 134.73 (s), 129.59 (s), 129.51 (s), 128.73 (s), 126.95 (s), 124.45 (s), 119.83 (s), 43.74 (s), 40.23 (s).

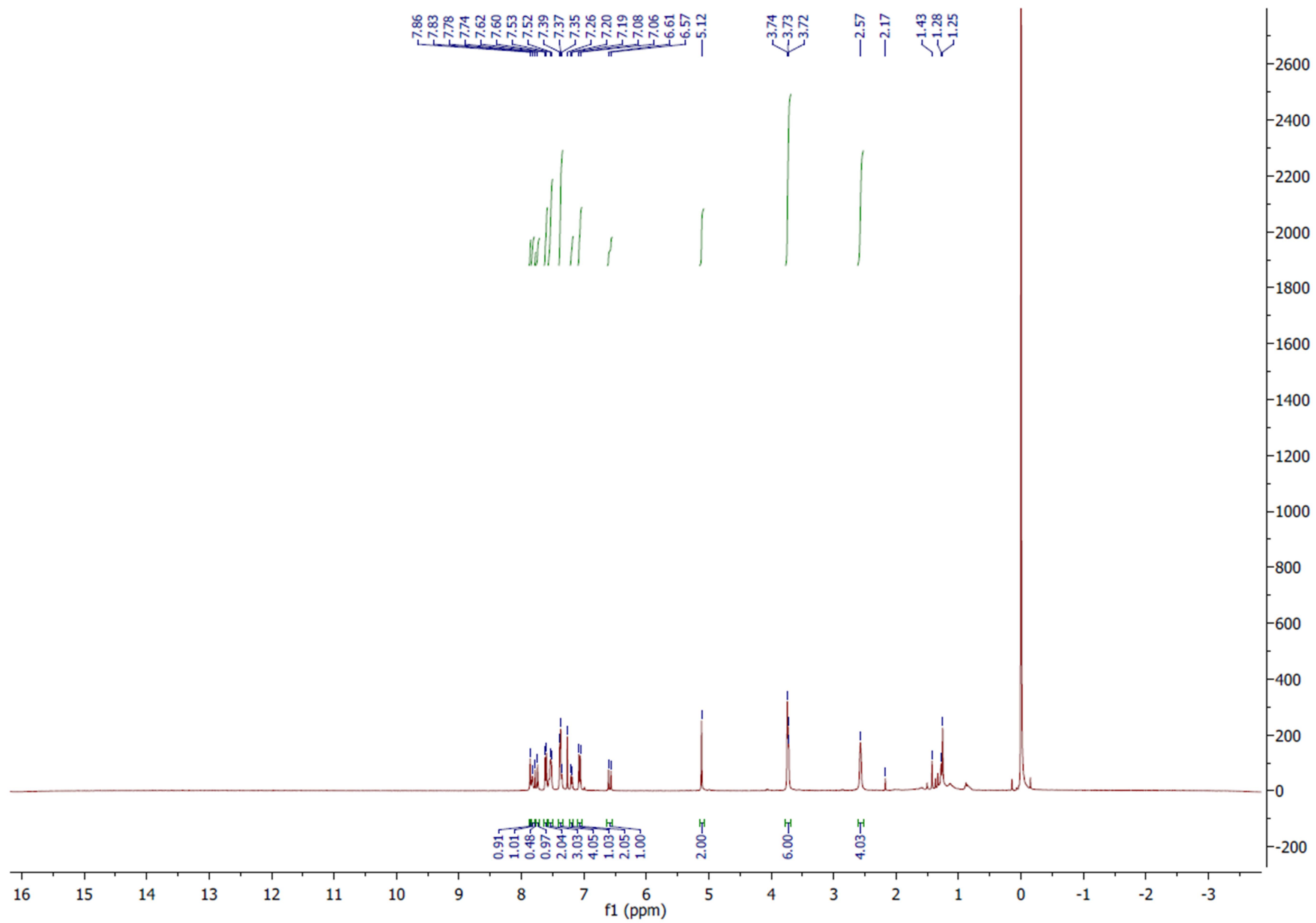
ND24_A	¹ H NMR (400 MHz, CDCl ₃) δ 7.33 – 7.24 (m, 2H), 7.21 (d, <i>J</i> = 7.1 Hz, 3H), 7.14 (s, 1H), 7.09 (s, 1H), 7.03 (t, <i>J</i> = 8.0 Hz, 1H), 6.58 (d, <i>J</i> = 7.7 Hz, 1H), 6.39 (d, <i>J</i> = 7.7 Hz, 1H), 3.57 (s, 2H), 3.02 (t, <i>J</i> = 7.6 Hz, 2H), 2.61 (t, <i>J</i> = 7.6 Hz, 2H).
ND24_B	¹ H NMR (400 MHz, CDCl ₃) δ 8.74 (d, <i>J</i> = 4.9 Hz, 2H), 8.28 (s, 1H), 7.93 (s, 1H), 7.65 (d, <i>J</i> = 5.5 Hz, 2H), 7.46 (s, 1H), 7.39 (s, 1H), 7.32 – 7.26 (m, 2H), 7.22 (dd, <i>J</i> = 10.9, 6.5 Hz, 4H), 7.11 (s, 1H), 3.01 (t, <i>J</i> = 7.6 Hz, 2H), 2.63 (t, <i>J</i> = 7.6 Hz, 2H).
ND24_C	¹ H NMR (400 MHz, CDCl ₃) δ 8.17 (d, <i>J</i> = 8.6 Hz, 2H), 7.51 (d, <i>J</i> = 8.6 Hz, 2H), 7.29 (d, <i>J</i> = 7.9 Hz, 2H), 7.25 – 7.17 (m, 3H), 7.15 – 6.93 (m, 3H), 6.50 (d, <i>J</i> = 7.8 Hz, 1H), 6.28 (d, <i>J</i> = 7.5 Hz, 1H), 4.45 (s, 2H), 4.30 (s, 1H), 3.02 (t, <i>J</i> = 7.6 Hz, 2H), 2.62 (t, <i>J</i> = 7.6 Hz, 2H).
ND24_D	¹ H NMR (400 MHz, CDCl ₃) δ 7.30 (d, <i>J</i> = 7.2 Hz, 1H), 7.25 – 6.86 (m, 9H), 6.66 (d, <i>J</i> = 8.0 Hz, 2H), 6.59 (d, <i>J</i> = 7.8 Hz, 1H), 6.36 (d, <i>J</i> = 7.9 Hz, 1H), 4.87 (s, 1H), 4.16 (s, 2H), 3.79 (s, 2H), 3.04 (t, <i>J</i> = 7.4 Hz, 2H), 2.62 (t, <i>J</i> = 7.4 Hz, 2H).



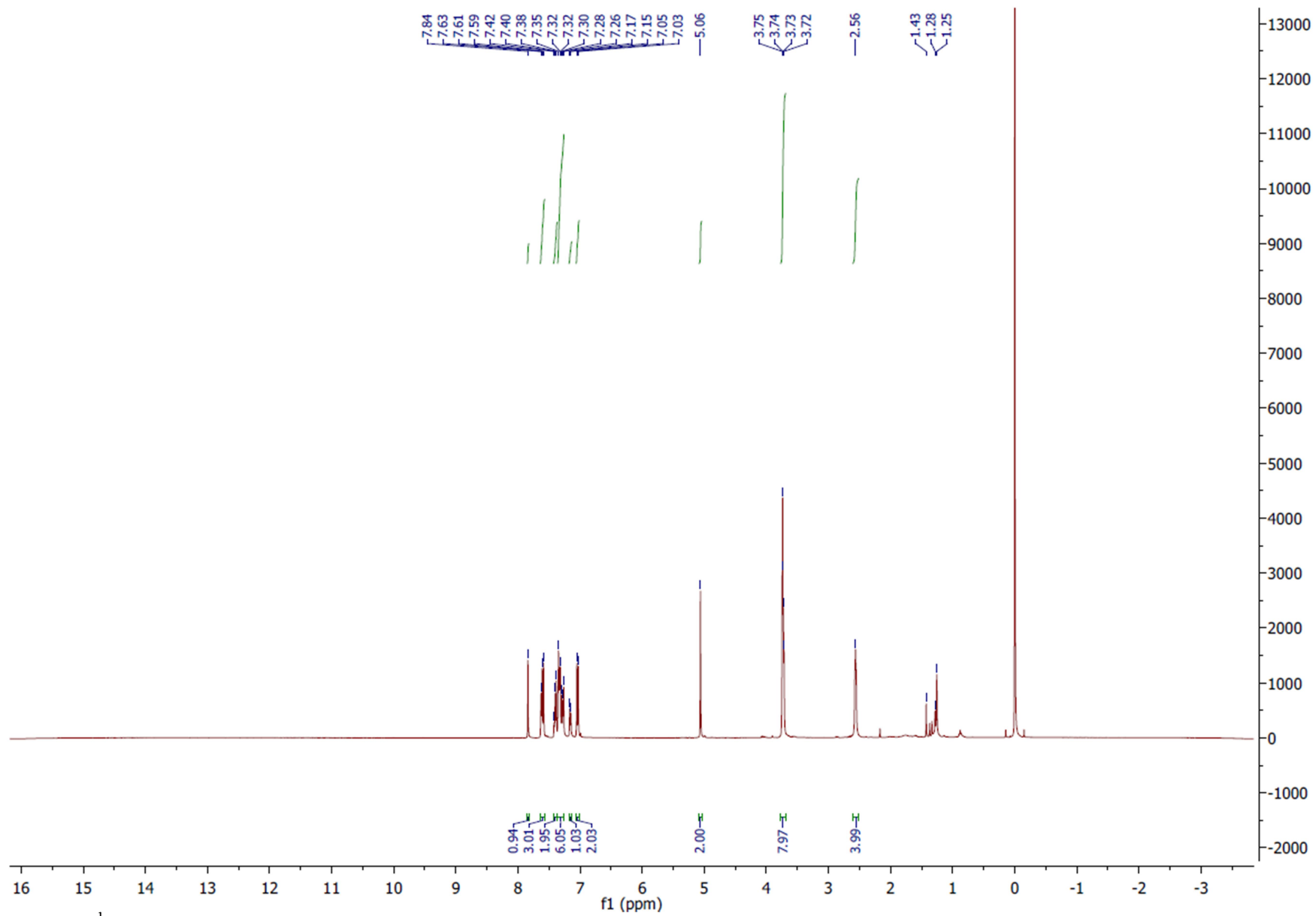
Slika P6. ^1H NMR spektar jedinjenja ND3_7.



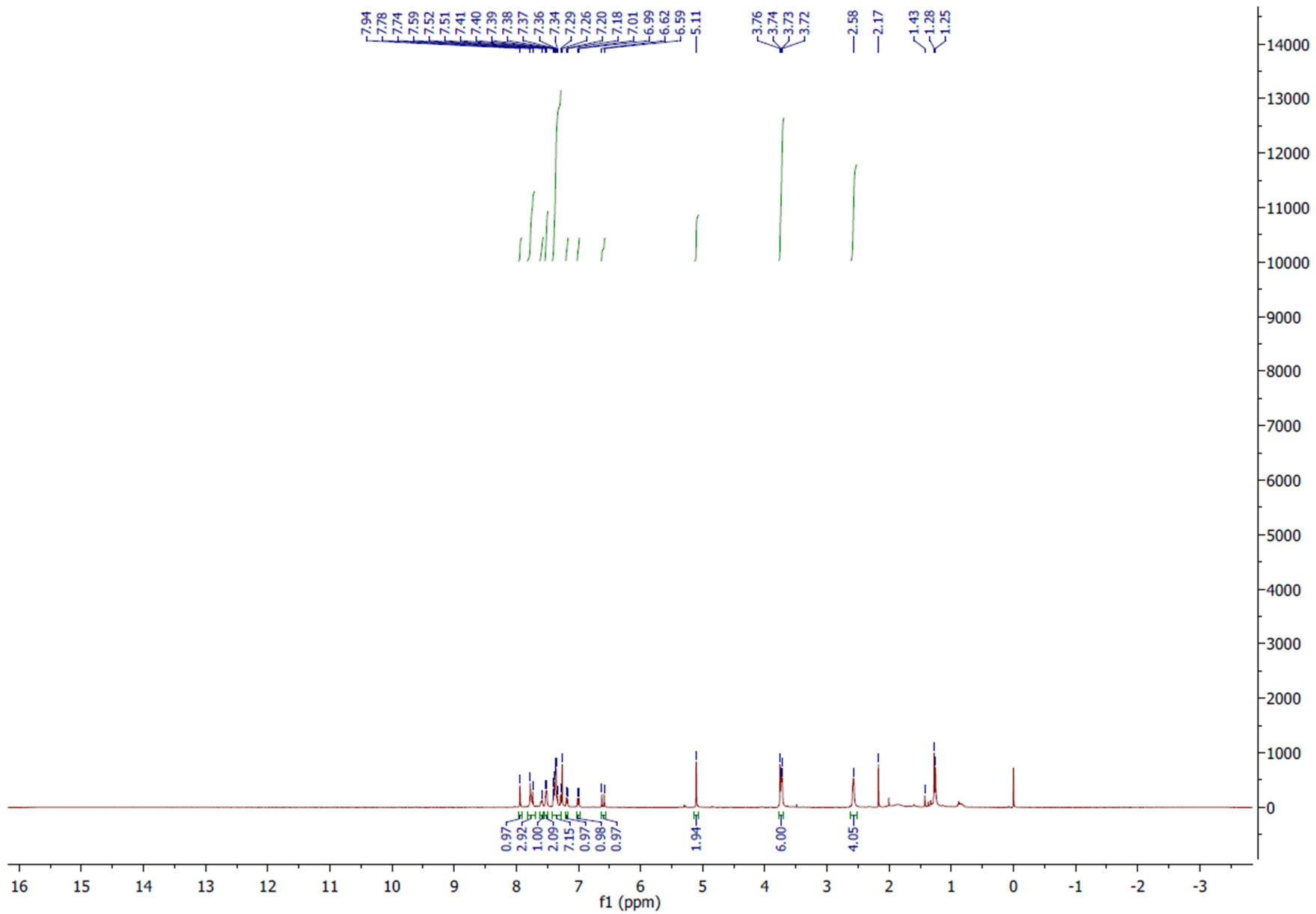
Slika P7. ^1H NMR spektar jedinjenja ND4_7.



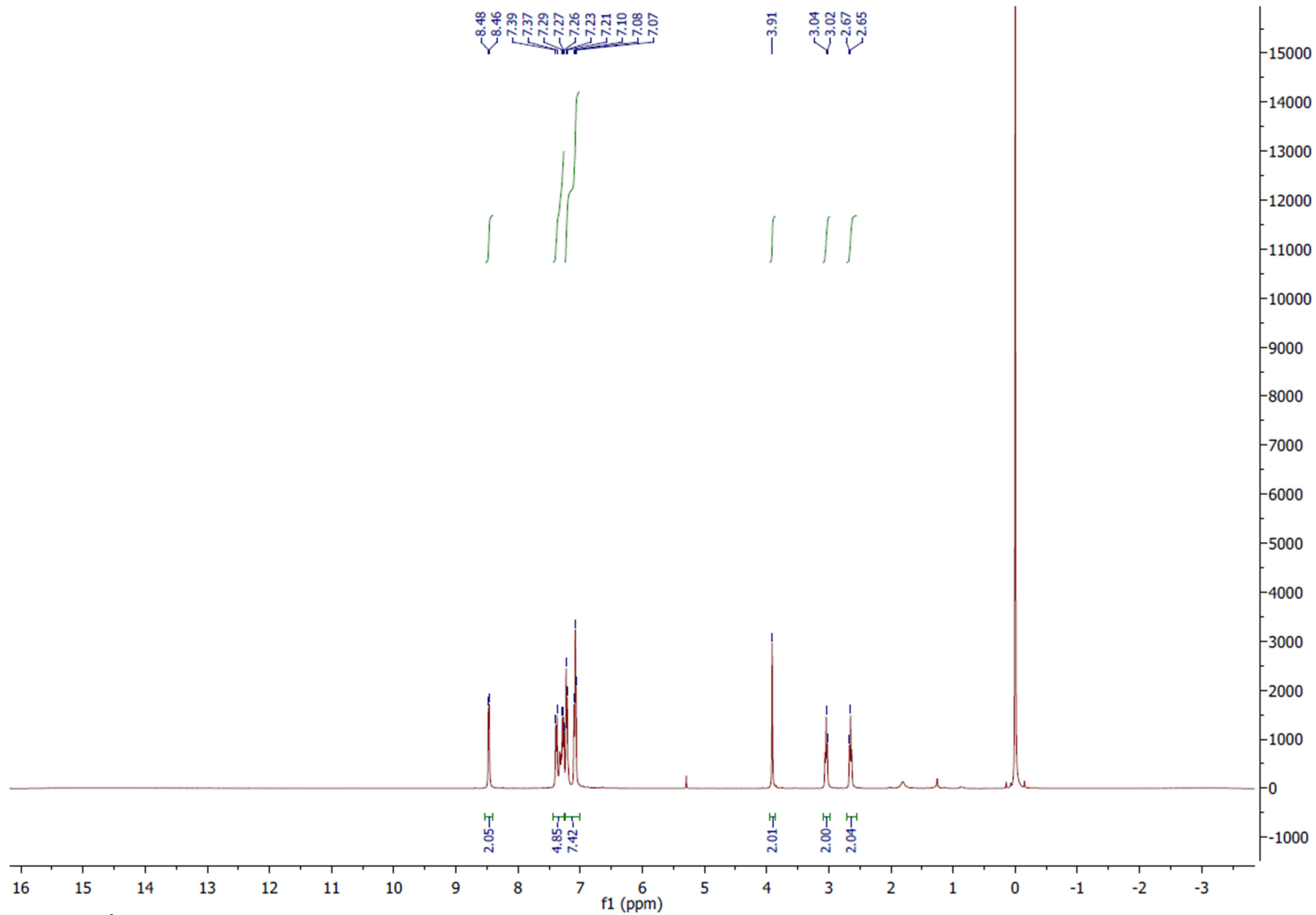
Slika P8. ^1H NMR spektar jedinjenja ND5_7.



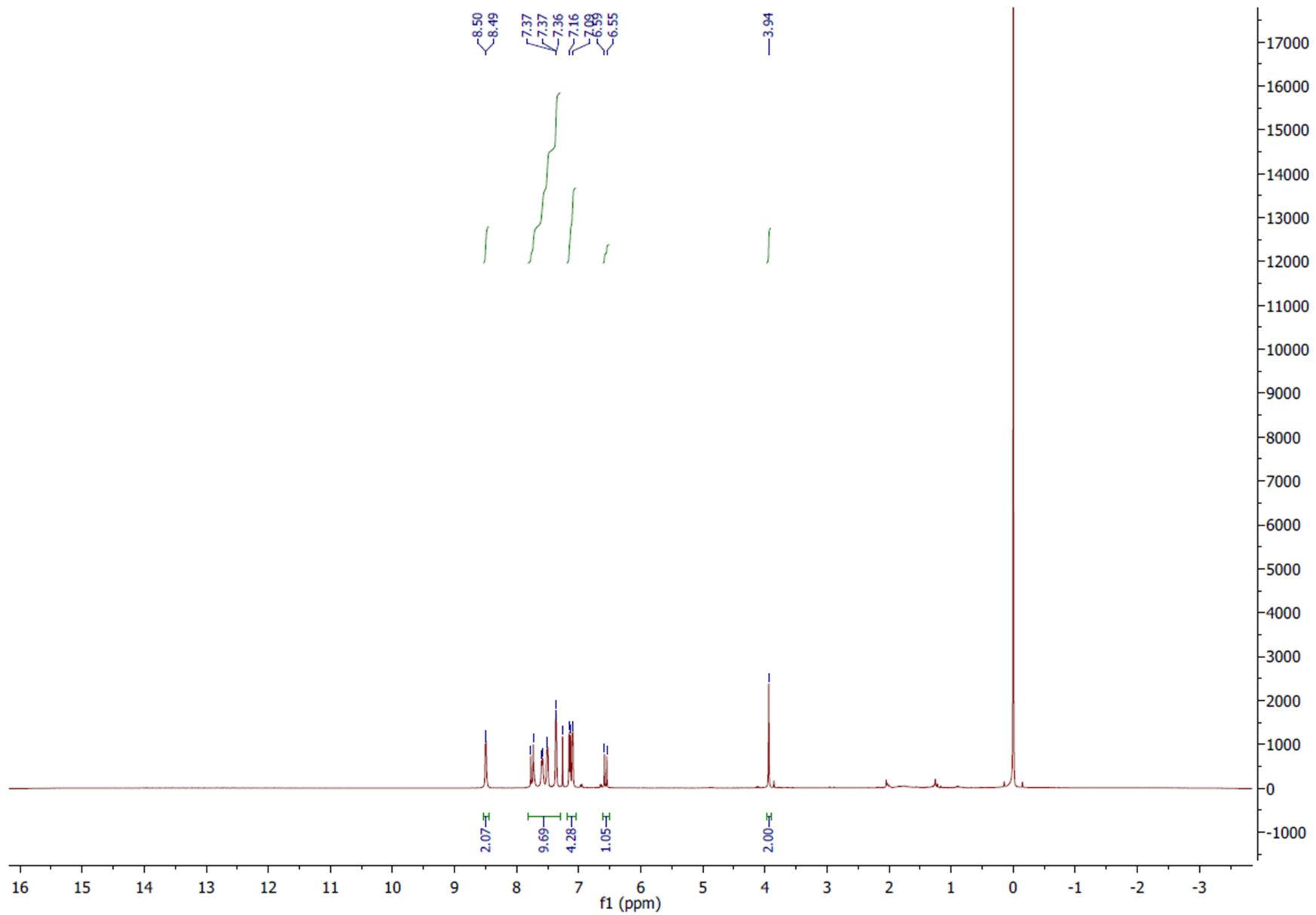
Slika P9. ¹H NMR spektar jedinjenja ND6_7.



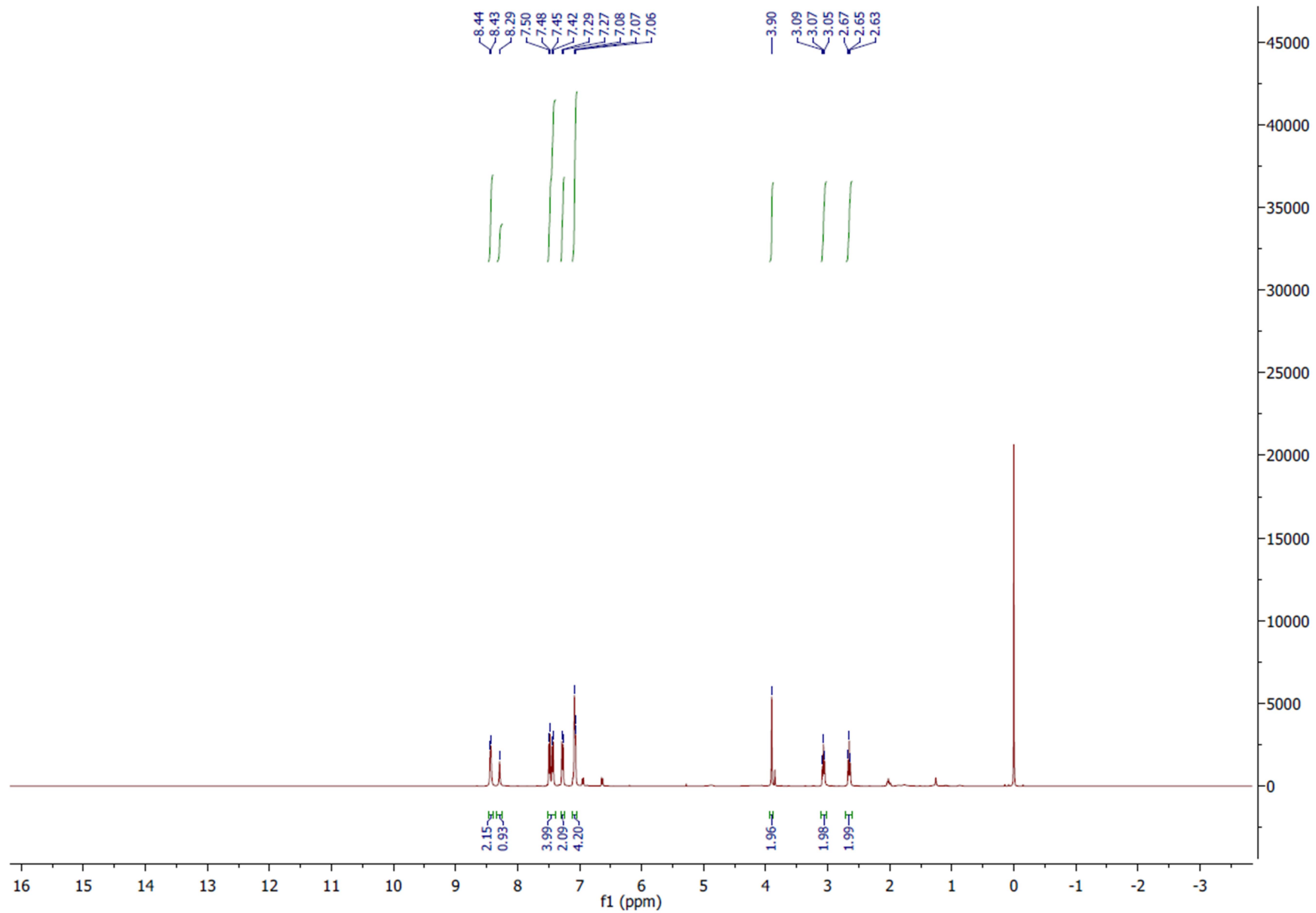
Slika P10. ^1H NMR spektar jedinjenja ND7_7.



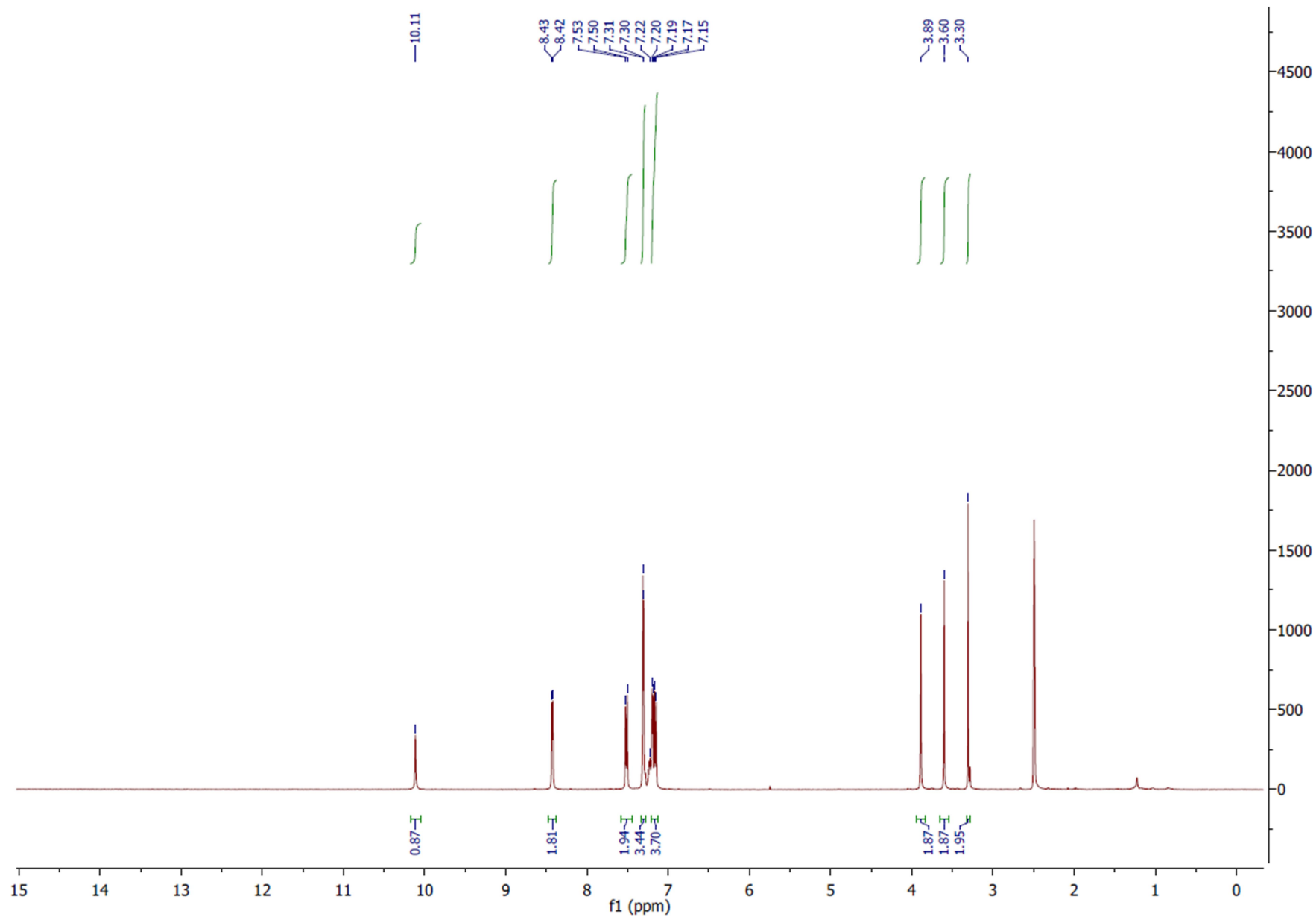
Slika P11. ¹H NMR spektar jedinjenja ND18.



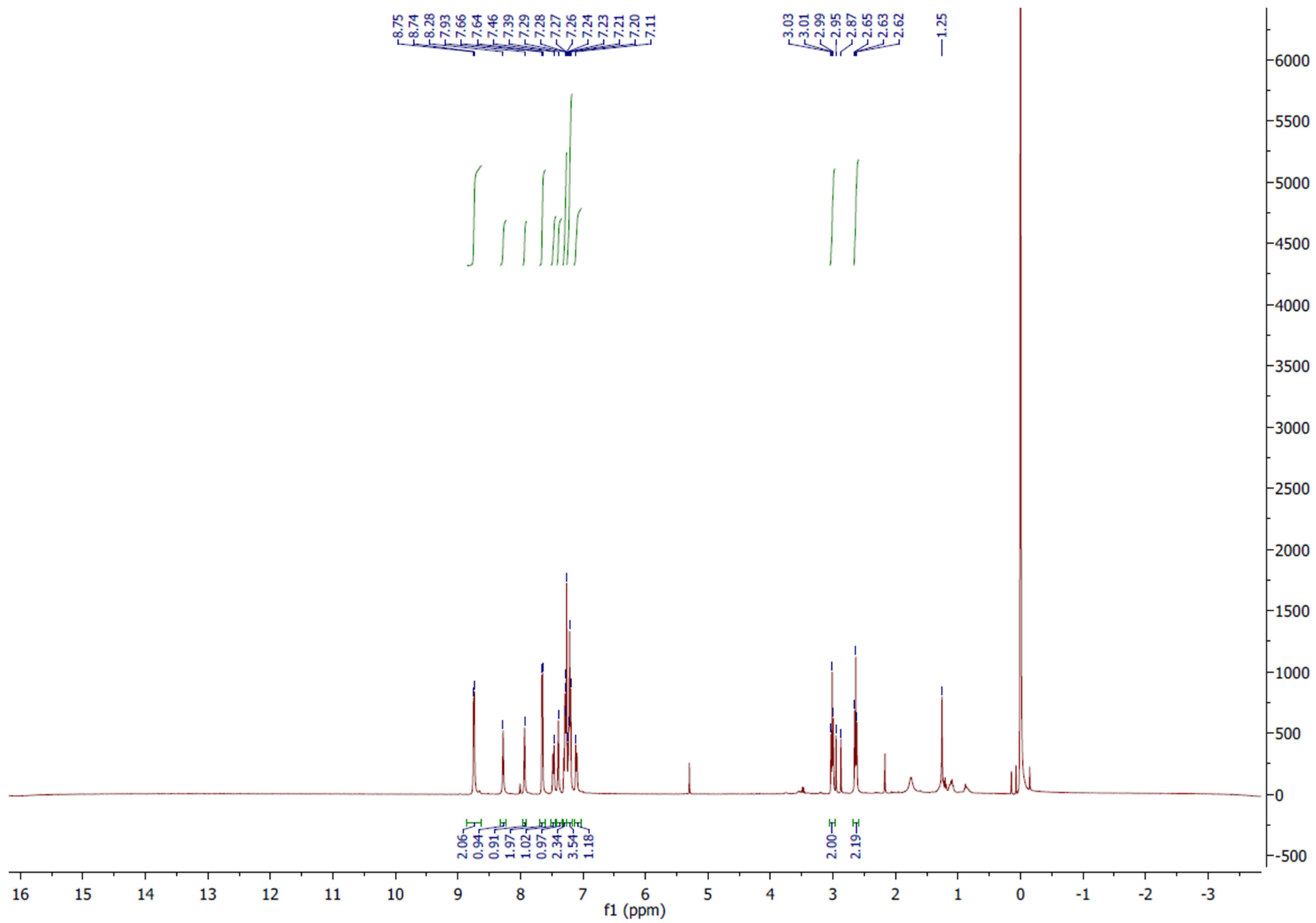
Slika P12. ^1H NMR spektar jedinjenja ND19.



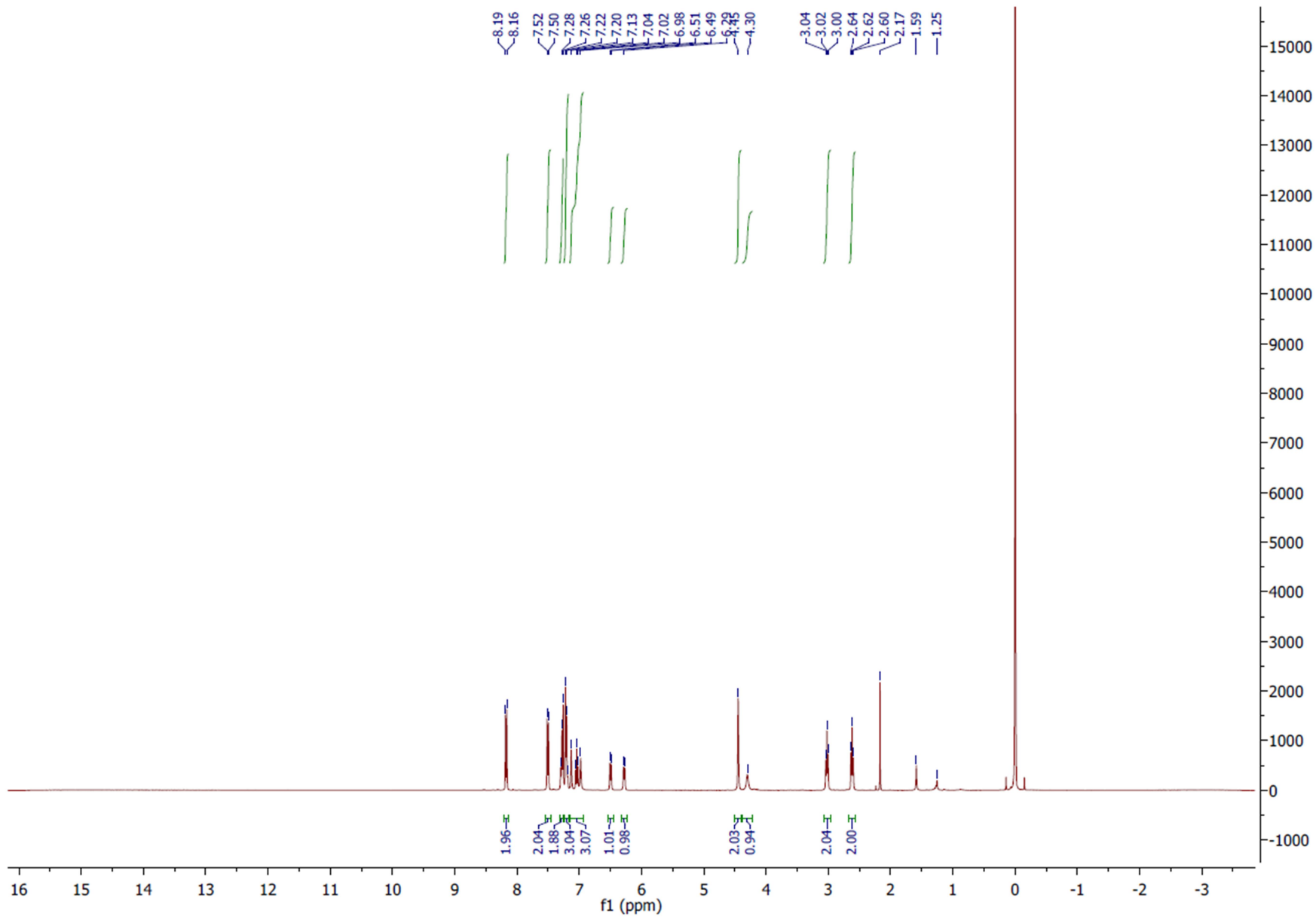
Slika P13. ^1H NMR spektar jedinjenja ND20.



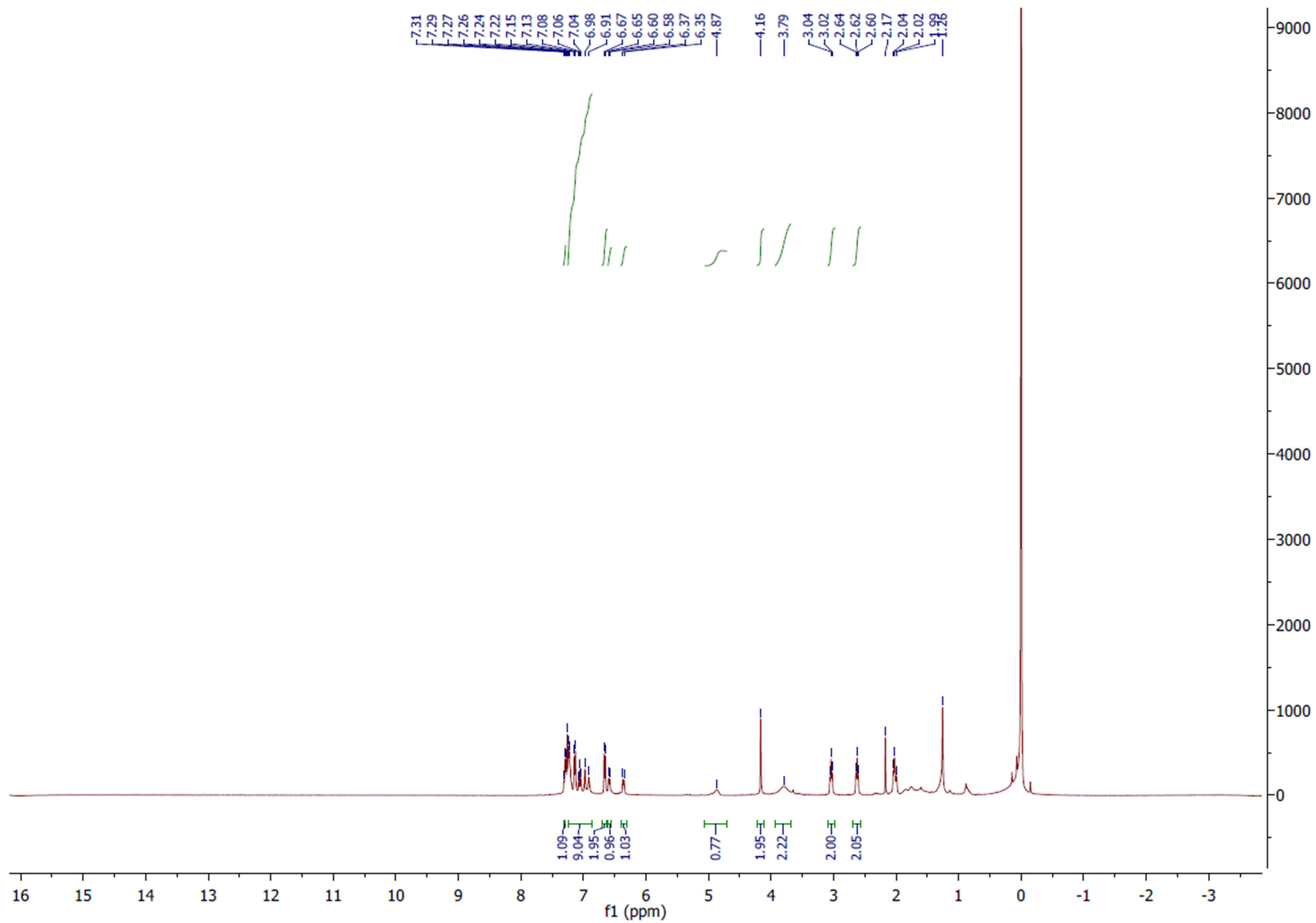
Slika P14. ¹H NMR spektar jedinjenja ND23.



Slika P15. ^1H NMR spektar jedinjenja ND24_B.

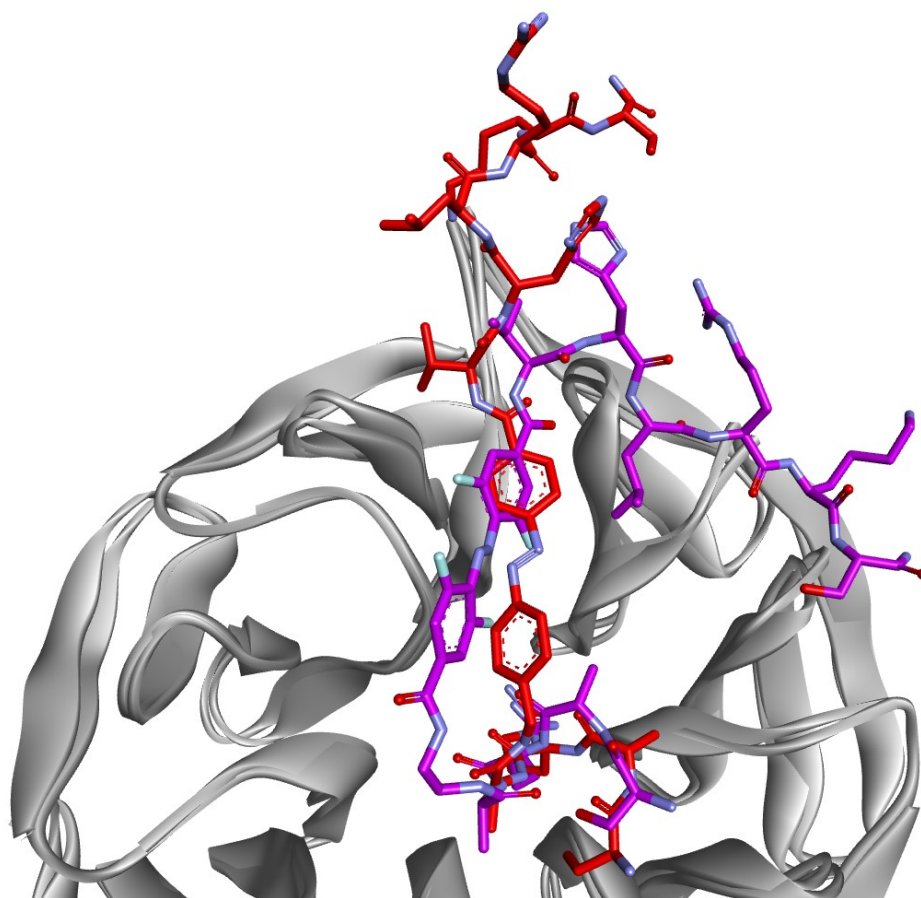


Slika P16. ^1H NMR spektar jedinjenja ND24_C.

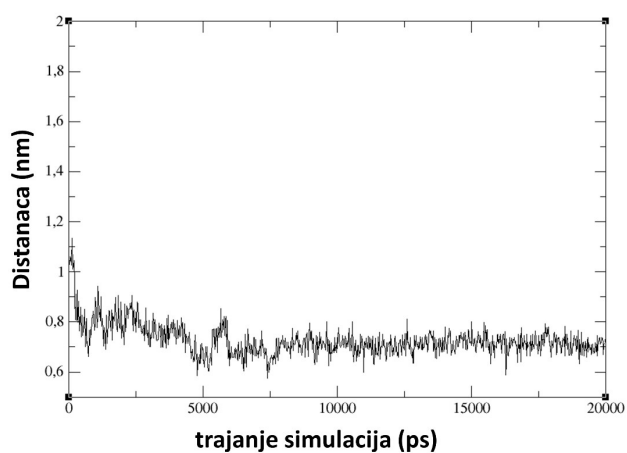


Slika P17. ^1H NMR spektar jedinjenja ND24_D.

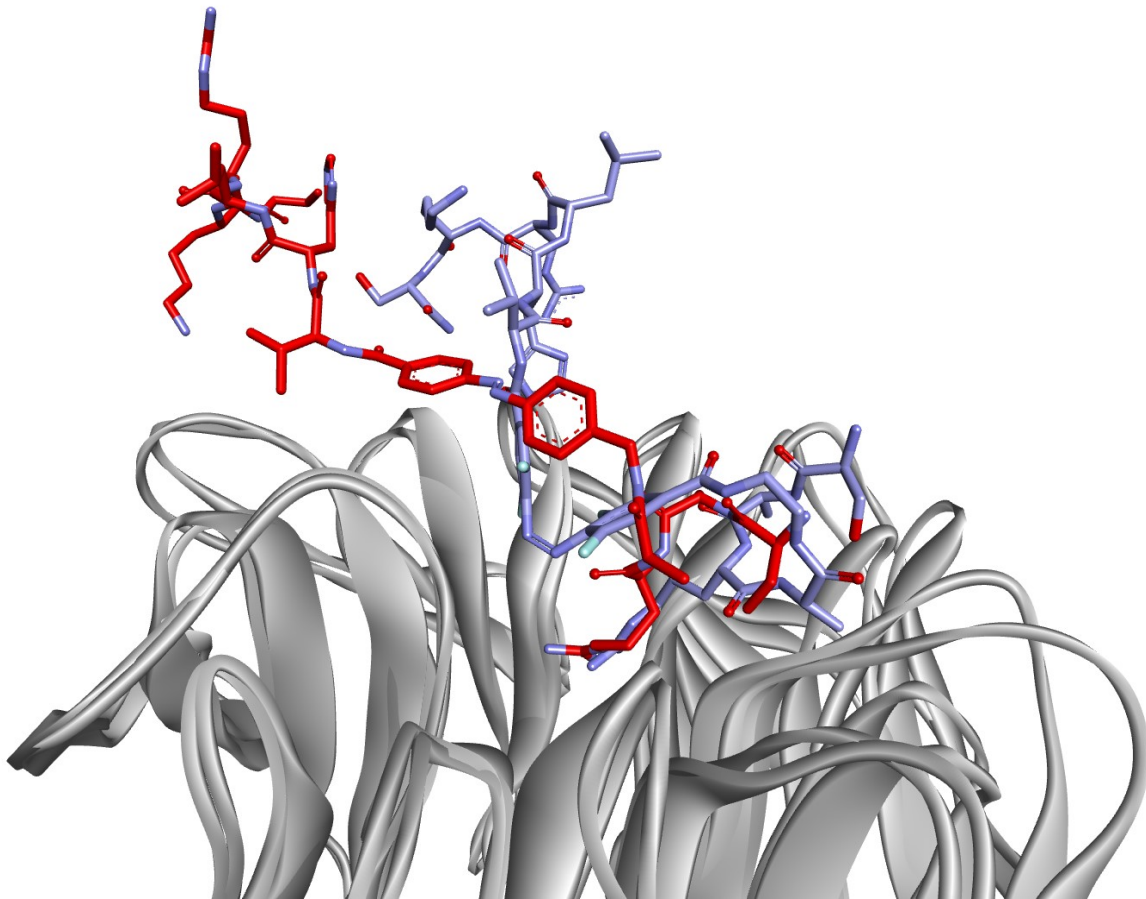
Prilog 7.11. Modelovane strukture kompleksa WDR5 sa lineranim fotoreaktivnim peptidomimeticima.



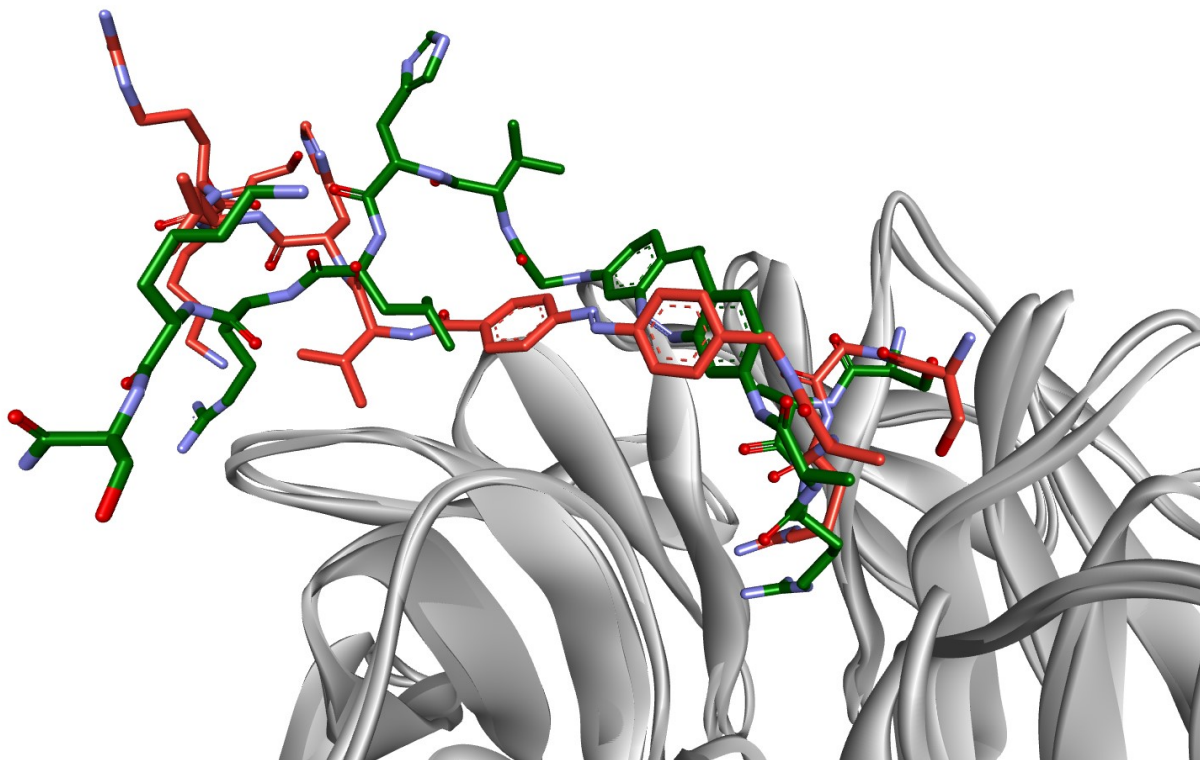
Slika P18. Poređenje između previđenih vezivnih modova *trans* AMBP (3) (crveno) i *trans* oF4Azo-peptida (5) (tamnoljubičasto).



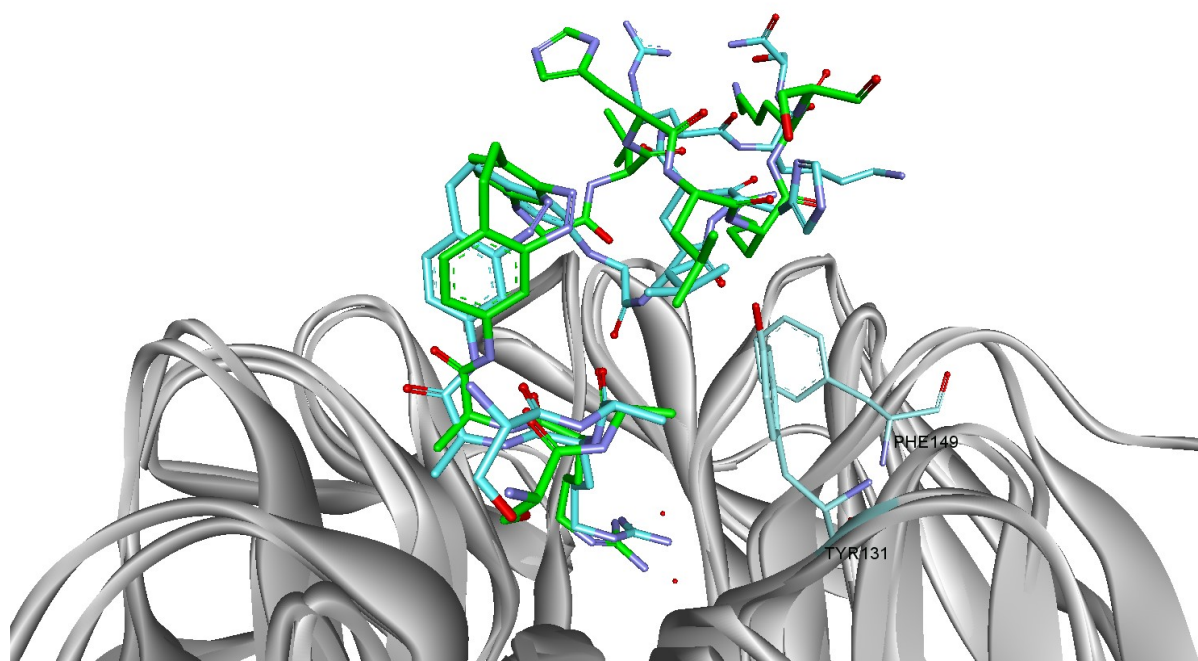
Slika P19. Distanca između leucina *trans* oF4Azo-peptida (5) (-VHLRKS sekvenca) i tetra-orto-fluoroazobenzena tokom 20 ns simulacija. Stabilizacija interakcije je postignuta nakon 7 ns.



Slika P20. Poređenje predviđenih vezivnih modova MD simulacija *trans* AMBP (3) (crveno) i *cis* oF4Azo-peptida (5) (svetlojubičasto).



Slika P21. Poređenje predviđenih vezivnih modova *trans* AMBP (3) (crveno) i *trans* cAzo-peptida (5) (zeleno).



Slika P22. Poređenje predviđenog vezivnog moda (svetloplava) i kristalografski razrešenog (zeleno) za *cis* cAzo-peptida (4).

Prilog 7.12. Rezultai analize dinamičkih mreža interakcija rezidua WDR5.

Tabela P14. Rezultati uzajamnog poređenja modela mreža između različitih sistema.

Sistemi A/B ^a		Sistemi C/D ^a	
Residua	D _n -Skor	Residua	D _n -Skor
His255	5.099489	Asp150	1.503232
His86	4.70898	Asp211	1.344799
His128	4.273141	Asp92 ^b	1.042917
Asp108 ^b	3.338209	Asp107 ^b	1.01655
Asn180	3.094606	Asn287	1.010103
Arg196 ^c	2.796166	Phe133 ^b	1.006474
Ser106 ^b	2.642996	Ser190 ^b	1.004777
Lys126	2.556442	Ser49 ^b	1.002467
Lys221 ^d	2.502327	Pro173 ^b	1.001611
Asp150	2.337324	Asn130 ^b	1.001167
His310	2.329052	Val48 ^b	1.0005
His170	2.285121	Leu194	0.572672
Asn265	1.916545	Asn265	0.559351
Pro311	1.824915	Arg196 ^c	0.549544
His178	1.703283	Asp192 ^c	0.542799
Asp92 ^b	1.680356	Glu313	0.536194
Glu292	1.65801	Ser91 ^b	0.524734
Lys162	1.631025	Asp66 ^b	0.517267
Lys52 ^b	1.628334	Ser64 ^b	0.515594
Ser175 ^b	1.592773	Ser148 ^b	0.514068
Leu167	1.575877	Ser148 ^b	0.514068
Leu111	1.571174		
Asn257	1.497048		
Lys239 ^d	1.48971		
Asp157	1.458518		
Ser148 ^b	1.310101		
Arg154	1.302673		
Thr237	1.261176		
Thr110	1.256864		
Trp156	1.234355		
Thr233	1.166917		
Tyr252 ^b	1.15467		
Leu194	1.145334		
Asp192 ^c	1.140995		
Glu58	1.121784		
Ile83	1.091371		
Ser190 ^{b,c}	1.091004		
Asp107 ^b	1.060644		
Pro173 ^b	1.0554		
Asn130 ^b	1.026998		

^a Usled većih razlika u dinamičkom ponašanju između sistema A i B, različite granične vrednosti su razmatrane. Za sisteme A i B, granična vrednost je iznosila 1, dok je za sistema C i D iznosila 0,5.

^b Rezidue unutar WIN vezivnog mesta.

^c Rezidue alternativnog komunikacionog puta između WIN i WBM mesta.

^d Rezidue WBM mesta.

Biografija

Nemanja B. Đoković je rođen 11. novembra 1992. godine u Čačku gde je završio osnovnu i srednju školu kao nosilac diploma „Vuk Karadžić“. Farmaceutski fakultet Univerziteta u Beogradu – smer magistar, farmacije upisao je 2011/2012. godine. Tokom studija bio je nagrađivan Godišnjom nagradom najboljim studentima integrisanih akademskih studija farmacije. Završni rad odbranio je 2016. godine na Katedri za farmaceutsku hemiju pod mentorstvom prof. dr Katarine Nikolić i time stekao titulu magistar farmacije (prosečna ocena 9,73). Doktorske akademske studije – modul Farmaceutska hemija, upisao je školske 2016/2017. Tokom doktorskih studija bio je nagrađivan Godišnjom nagradom Farmaceutskog fakulteta za najbolje naučnoistraživačke radove studenata poslediplomskih studija i sa dve nagrade za izuzetan doprinos naučno-istraživačkom radu fakulteta. Tokom izrade doktorske disertacije (2017-2019. godine) usavršavao se u Laboratoriji za istraživanje sirtuina Univerziteta Istočne Finske, Kuopio, Finska (mentor dr Maija Lahtela Kakkonen). Od oktobra 2016. do marta 2017. bio je zaposlen u svojstvu stručnog saradnika na Katedri za farmaceutsku hemiju, a tokom aprila 2017. na Katedri za analitičku hemiju. Od oktobra 2017. godine zaposlen je u svojstvu istraživača-pripravnika, zatim i istraživača-saradnika na projektu Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije. Od juna 2022. zaposlen je kao asistent na Katedri za farmaceutsku hemiju. U okviru naučno-istraživačkog rada, Nemanja Đoković je objavio 16 radova i jedno poglavlje u monografiji od međunarodnog značaja i to: 3 rada u međunarodnom časopisu izuzetnih vrednosti (M21a), 7 radova u vrhunskim međunarodnim časopisima (M21), 5 radova u istaknutim međunarodnim časopisima (M22) i jedan rad u vrhunskom nacionalnom časopisu (M51). Na međunarodnim kongresima ima 19 saopštenja štampanih u izvodu.

Spisak radova koji čine deo doktorske disertacije

- 1. Đoković, N.;** Djuric, A.; Ruzic, D.; Srdic-Rajic, T.; Nikolic, K. Correlating Basal Gene Expression across Chemical Sensitivity Data to Screen for Novel Synergistic Interactors of HDAC Inhibitors in Pancreatic Carcinoma. *Pharmaceuticals* 2023, 16 (2), 294. <https://doi.org/10.3390/ph16020294>. (M21, Chemistry, Medicinal (16/63), IF₂₀₂₁=5,215).
- 2. Đoković, N.;** Rahnasto-Rilla, M.; Lougiakis, N.; Lahtela-Kakkonen, M.; Nikolic, K. SIRT2i_Predictor: A Machine Learning-Based Tool to Facilitate the Discovery of Novel SIRT2 Inhibitors. *Pharmaceuticals* 2023, 16 (1), 127. <https://doi.org/10.3390/ph16010127>. (M21, Chemistry, Medicinal (16/63), IF₂₀₂₁=5,215).
- 3. Đoković, N.;** Ruzic, D.; Rahnasto-Rilla, M.; Srdic-Rajic, T.; Lahtela-Kakkonen, M.; Nikolic, K. Expanding the Accessible Chemical Space of SIRT2 Inhibitors through Exploration of Binding Pocket Dynamics. *J. Chem. Inf. Model.* 2022, 62 (10), 2571–2585. <https://doi.org/10.1021/acs.jcim.2c00241>. (M21, Chemistry, Medicinal (9/63), IF₂₀₂₁=6,162).
- Albert, L.; Nagpal, J.; Steinchen, W.; Zhang, L.; Werel, L.; **Đoković, N.;** Ruzic, D.; Hoffarth, M.; Xu, J.; Kaspereit, J.; Abendroth, F.; Royant, A.; Bange, G.; Nikolic, K.; Ryu, S.; Dou, Y.; Essen, L.-O.; Vázquez, O. Bistable Photoswitch Allows in Vivo Control of Hematopoiesis. *ACS Cent. Sci.* 2022, 8 (1), 57–66. <https://doi.org/10.1021/acscentsci.1c00434>. (deljeno treće autorsvo, M21a, Chemistry, Multidisciplinary (12/180), IF₂₀₂₁=18,728).
- Albert, L.; Peñalver, A.; **Đoković, N.;** Werel, L.; Hoffarth, M.; Ruzic, D.; Xu, J.; Essen, L.-O.; Nikolic, K.; Dou, Y.; Vázquez, O. Modulating Protein–Protein Interactions with Visible-Light-Responsive Peptide Backbone Photoswitches. *ChemBioChem* 2019, 20 (11), 1417–1429. <https://doi.org/10.1002/cbic.201800737>. (M22, Chemistry/Medicinal (30/61) IF₂₀₁₈ = 2.641).

Izjava o autorstvu

Ime i prezime autora: Nemanja Đoković

Broj indeksa: 17/2016

Izjavljujem

da je doktorska disertacija pod naslovom

Simulacije molekulske dinamike, dizajn, sinteza i *in vitro* ispitivanje inhibitora epigenetskih proteina kao potencijalnih antineoplastika

- rezultat sopstvenog istraživačkog rada;
- da disertacija u celini ni u delovima nije bila predložena za sticanje druge diplome prema studijskim programima drugih visokoškolskih ustanova;
- da su rezultati korektno navedeni i
- da nisam kršio/la autorska prava i koristio/la intelektualnu svojinu drugih lica.

Potpis autora

U Beogradu, _____

Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Ime i prezime autora	<u>Nemanja Đoković</u>
Broj indeksa	<u>17/2016</u>
Studijski program	<u>Doktorske akademske studije – farmaceutska hemija</u>
Naslov rada	<u>Simulacije molekulske dinamike, dizajn, sinteza i <i>in vitro</i> ispitivanje inhibitora epigenetskih proteina kao potencijalnih antineoplastika</u>
Mentor	<u>Prof. dr Katarina Nikolić, redovni profesor na Katedri za farmaceutsku hemiju, Farmaceutski fakultet, Univerzitet u Beogradu</u>

Izjavljujem da je štampana verzija mog doktorskog rada istovetna elektronskoj verziji koju sam predao radi pohranjivanja u **Digitalnom repozitorijumu Univerziteta u Beogradu**.

Dozvoljavam da se objave moji lični podaci u vezi sa dobijanjem akademskog naziva doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada.

Ovi lični podaci mogu se objaviti na mrežnim stranicama digitalne biblioteke, u elektronskom katalogu i u publikacijama Univerziteta u Beogradu.

Potpis autora

U Beogradu, _____

Izjava o korišćenju

Ovlašćujem Univerzitetsku biblioteku „Svetozar Marković“ da u Digitalni repozitorijum Univerziteta u Beogradu unese moju doktorsku disertaciju pod naslovom:

Simulacije molekulske dinamike, dizajn, sinteza i *in vitro* ispitivanje inhibitora epigenetskih proteina kao potencijalnih antineoplastika

koja je moje autorsko delo.

Disertaciju sa svim priložima predao sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalnom repozitorijumu Univerziteta u Beogradu i dostupnu u otvorenom pristupu mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio.

1. Autorstvo (CC BY)
2. Autorstvo – nekomercijalno (CC BY-NC)
3. Autorstvo – nekomercijalno – bez prerada (CC BY-NC-ND)
4. Autorstvo – nekomercijalno – deliti pod istim uslovima (CC BY-NC-SA)
5. Autorstvo – bez prerada (CC BY-ND)
6. Autorstvo – deliti pod istim uslovima (CC BY-SA)

(Molimo da zaokružite samo jednu od šest ponuđenih licenci.
Kratak opis licenci je sastavni deo ove izjave).

Potpis autora

U Beogradu, _____

1. **Autorstvo.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence, čak i u komercijalne svrhe. Ovo je najslobodnija od svih licenci.
2. **Autorstvo – nekomercijalno.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence. Ova licenca ne dozvoljava komercijalnu upotrebu dela.
3. **Autorstvo – nekomercijalno – bez prerada.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, bez promena, preoblikovanja ili upotrebe dela u svom delu, ako se navede ime autora na način određen od strane autora ili davaoca licence. Ova licenca ne dozvoljava komercijalnu upotrebu dela. U odnosu na sve ostale licence, ovom licencom se ograničava najveći obim prava korišćenja dela.
4. **Autorstvo – nekomercijalno – deliti pod istim uslovima.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence i ako se prerada distribuira pod istom ili sličnom licencom. Ova licenca ne dozvoljava komercijalnu upotrebu dela i prerada.
5. **Autorstvo – bez prerada.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, bez promena, preoblikovanja ili upotrebe dela u svom delu, ako se navede ime autora na način određen od strane autora ili davaoca licence. Ova licenca dozvoljava komercijalnu upotrebu dela.
6. **Autorstvo – deliti pod istim uslovima.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence i ako se prerada distribuira pod istom ili sličnom licencom. Ova licenca dozvoljava komercijalnu upotrebu dela i prerada. Slična je softverskim licencama, odnosno licencama otvorenog koda.